User Modeling and User-Adapted Interaction (2023) 33:967–1014
https://doi.org/10.1007/s11257-022-09353-8

Enhancing user awareness on inferences obtained from fitness trackers data

Alexia Dini Kounoudes[1] · Georgia M. Kapitsaki[1] · Ioannis Katakis[2]

Received: 5 May 2022 / Accepted in revised form: 21 December 2022 / Published online: 17 January 2023

Abstract

In the IoT era, sensitive and non-sensitive data are recorded and transmitted to multiple service providers and IoT platforms, aiming to improve the quality of our lives through the provision of high-quality services. However, in some cases these data may become available to interested third parties, who can analyse them with the intention to derive further knowledge and generate new insights about the users, that they can ultimately use for their own benefit. This predicament raises a crucial issue regarding the privacy of the users and their awareness on how their personal data are shared and potentially used. The immense increase in fitness trackers use has further increased the amount of user data generated, processed and possibly shared or sold to third parties, enabling the extraction of further insights about the users. In this work, we investigate if the analysis and exploitation of the data collected by fitness trackers can lead to the extraction of inferences about the owners routines, health status or other sensitive information. Based on the results, we utilise the PrivacyEnhAction privacy tool, a web application we implemented in a previous work through which the users can analyse data collected from their IoT devices, to educate the users about the possible risks and to enable them to set their user privacy preferences on their fitness trackers accordingly, contributing to the personalisation of the provided services, in respect of their personal data.

Keywords Internet of things · Fitness trackers · User awareness · User-centred privacy · Personalised services · Privacy preferences

✉ Alexia Dini Kounoudes
adini-01@ucy.ac.cy

Georgia M. Kapitsaki
gkapi@ucy.ac.cy

Ioannis Katakis
katakis.i@unic.ac.cy

[1] Computer Science Department, University of Cyprus, 1 University Avenue, 2109 Nicosia, Cyprus

[2] Department of Computer Science, School of Sciences and Engineering, University of Nicosia, 2417 Nicosia, Cyprus

1 Introduction

With the ability to connect and control billions of devices and get access to valuable data, the Internet of things (IoT) is shaping the future of technology and society, as it is estimated that the number of connected devices will rise to 50 billion by 2030 (Vailshery 2021). The popularity of IoT devices, such as smart home devices and fitness trackers, has boosted the acquisition, exchange and distribution of data generated by their users. The amount of data shared between IoT devices is prodigious, as more than 150 zettabytes (150 trillion gigabytes) of data will be generated by 2025 (IotaComm 2020). Furthermore, "it is estimated that today the average person creates 1.5 GB of data on average daily" (Krzanich 2016). It is no wonder that today the phrase "data is the new gold" (Forbes 2019; WEF 2020; CEOToday 2020) is a metaphor describing a new paradigm that revolutionises the world.

In an interview in 2000 (Dennedy et al. 2014), the late Andrew Grove, CEO of Intel Corporation, prophetically stated that *"privacy is one of the biggest problems in this new electronic age"*. The protection of personal data forms a principal citizen right safeguarded in the European Union that is particularly pertinent in the IoT domain. The EU General Data Protection Regulation (GDPR),[1] introduced in 2018, aims to make the protection of this right effective by providing a high level of data protection. Furthermore, the regulation intents to provide a generic framework for the protection of the privacy of the users and their personal data and to provide awareness to the users of how their data are collected and processed. The GDPR is highly concerned with fitness trackers, as their functionality involves the use of personal data, and as such they have to comply with its directions. Since transparency is key in the application of GDPR, it is essential that fitness trackers users become aware about how their personal data are processed (Becher et al. 2020). These devices are used to monitor the users' daily fitness and physical activities and they collect enormous amounts of highly sensitive personalised body, health and fitness data, like activity, steps count, temperature, sleep patterns or calories burnt, using embedded sensors such as pedometers, accelerometers, GPS, heart rate monitors and altimeters (Yan et al. 2015).

In the existing literature, limited attention has been given to the development of user awareness mechanisms that can assist the users in understanding how the data created by their smart devices can be exploited for the extraction of inferences regarding their daily activities and lifestyle in general. There is an imperative need for the development of such tools, as fitness trackers collect sensitive personal information that can be acquired by unauthorised third parties without user awareness (Kounoudes and Kapitsaki 2020), and also because these devices have become the perfect prey for attacks and data breaches, due to the lack of strict security guidelines and the sensitive nature of the data collected by them (Masuch et al. 2021). At present, existing awareness mechanisms come in the form of tedious privacy policies (Alqhatani and Lipford 2021) that the users generally tend to ignore; thus, further research is required in order to design the necessary tools and approaches to make the users aware of how their smart devices data can be exploited by third parties presenting the information in a direct and comprehensive way (Kröger 2018) and enable them to assimilate how

[1] https://gdpr.eu/.

to reduce these risks, by suggesting simple solutions as, for example, by altering their privacy preferences.

This is the focus of this work, where by concentrating on fitness trackers from three brands, namely Fitbit, Garmin and Xiaomi, we investigate if the analysis and exploitation of the data collected by those trackers can lead to the extraction of inferences about the owners routines, health status or other sensitive information. We utilise a data inference framework introduced in our previous work (Kounoudes et al. 2021), where by using a number of machine learning, statistical analysis and modelling techniques we aim to verify that such inferences are possible in order to raise user awareness about them. These techniques are applied in the PrivacyEnhAction privacy tool introduced in our previous work, a web application through which the users can analyse data collected from their smart devices (smart water meters or motion sensors) with the objective to be informed about potential privacy vulnerabilities and possible inferences that emerge from the use of these devices, and thereupon to be able to change and set their user privacy preferences on their devices appropriately, contributing in this way to the personalisation of the provided services, in connection with their personal data. The tool has now been extended to include three fitness trackers brands in the list of smart devices whose data can be analysed for inference detection by the users. To that end, our work is user-oriented aiming to raise user awareness regarding privacy in the area of IoT and this dimension makes it distinct from other works in the area.
The main contributions of this work are:

- We lay out a list of possible inferences that pose a threat to user privacy when using fitness trackers, and we attempt to identify which specific inferences can be drawn from specific data collected from Fitbit, Garmin and Xiaomi fitness trackers.
- We present an implementation of these functionalities in PrivacyEnhAction aiming to increase user awareness in relation to privacy when using fitness trackers.
- We provide the results of our research consisting of two questionnaires targeting fitness trackers users aiming to evaluate if their interaction with the PrivacyEnhAction application has increased their awareness.

The rest of the paper is organised as follows: In Sect. 2, we provide an overview of the related work as well as some background knowledge utilised in the rest of the paper. In Sect. 3, we present the methodology we used in this work. In Sect. 4, we present a review of the privacy policies of the fitness trackers under study. In Sect. 5, we provide an analysis of the possible inferences that can be drawn about a user from fitness trackers data based on the available literature. In Sect. 6, we describe the methodology used to collect, examine and analyse the data in the three fitness trackers scenarios under study, and we explain how the inference detection analysis takes place in each case. Section 7 provides details about the implementation of these new functionalities in the PrivacyEnhAction application, while in Sect. 8 we present the results of the user evaluation process. Section 9 discusses the findings and gives an overview of the limitations of this work, while Sect. 10 concludes this study.

2 Background

In this section, we present research in the fields of privacy protection in wearables, inference extraction from fitness trackers data and also related works in the area that engage in the collection and analysis of the opinions and perceptions of fitness trackers users in relevance to the protection of their privacy.

2.1 Privacy protection in wearables

Fitness wearables include devices like sport watches, smartwatches, wristbands, chest straps and other smart gear that monitor and track the number of steps we take every day, how many stairs we climb, the number of hours we sleep every night or the quality of our sleep, among others. Studies have shown that smartphone users are most likely to own a fitness wearable (Balas et al. 2020), while compatible Fitbit devices enable the users to make contactless payments, providing additional services. Data collected by wearables can be exploited in the pursue of inferring information regarding bodily activities like walking or running (Chen and Shen 2017), while smartwatch data have been successfully used for the recognition of user eating activities (Thomaz et al. 2015), drinking activities (Parate 2014) or smoking (Tang 2014).

Since the essence of wearables and fitness trackers does not usually allow a high level of interaction between the device and the users, a user interface is proposed by Mohzary et al. (2020) for capturing the privacy preferences of the users in each application they use. The presented GUI aims to educate the user about data access requests and protect her personal data. The privacy vulnerabilities and threats of using fitness trackers, in particular the Fitbit smartwatch, are explored in another work (Blow et al. 2020), by analysing the device features and potential security risks. The authors present a list of actions to diminish these vulnerabilities and they propose a number of best practices for wearables manufacturers to provide balance between functionality and privacy protection.

As the sensitive information collected by fitness trackers needs to be protected, a method for accumulating and processing health data in a privacy-preserving way is presented by Kim et al. (2020). Local differential privacy is being used adopting a sampling-based data collection scheme that accomplishes an important advancement in accuracy than simpler solutions, providing better privacy protection on the data collected. An anonymisation approach is proposed by Arca and Hewett (2020) to protect the privacy of the users data from smart health devices, by generalising pivotal data aiming to make it arduous to re-identify a user. According to the authors, the results of their technique demonstrate that with a small compromise on computational cost and data retention, the solution is effective for privacy protection. An analysis of the third parties that communicate with fitness trackers and their associated smartphone applications is presented by Kazlouski et al. (2020), where any unexpected—from the privacy point of view—third parties are identified. The aim of this work is to urge the users to study the privacy policies of devices before purchasing them to learn more about what personal data are being shared.

While Psychoula et al. (2020) were occupied with user privacy awareness in the area of wearables and IoT services by presenting a framework that could be used as guidance to developers and service providers in order to integrate privacy risk user awareness in their products, no other work to the best of our knowledge has been involved with raising user awareness in relation to the inferences that can be extracted about the users from their fitness trackers data.

2.2 Information inference as a privacy threat in IoT

In the literature, it has been shown that seemingly harmless data from smart devices can be used to infer eminently personal information about the users (Kröger 2018). Machine learning techniques and big data analytics have been used for drawing vigorous inferences from apparently harmless data or identified behaviour, compromising a basic privacy law, which is to allow a person to control who knows what about them (Horvitz and Mulligan 2015). Similar techniques are also used for making predictions about people's private lives, behaviours, habits and preferences, establishing the perfect conditions for discrimination, prejudicial and intrusive decision-making against the people involved (Wachter and Mittelstadt 2019), creating a crucial threat to user privacy. Recently, these privacy-related concerns have expanded from personal worries to social issues, as "anonymised" fitness tracking data from Strava, a widely used application for tracking activity and exercise, were released in the form of an "anonymised" heat map. The company mapped its accumulated activity data of two years in order to display the most visited areas in the map. However, US secret war zone locations and military bases were highlighted as soldiers habitually upload their fitness tracking data to Strava, creating a massive security threat as sensitive government and military sites were exposed (Whittaker 2018).

In the domain of IoT, inferences are personal information that are not consciously provided by the users themselves, but extracted by data controllers or other third parties from given data. This is a common approach in the area of machine learning; still inferences can be obtained without the use of advanced techniques. A "current" example of an inference that can be extracted without the use of machine learning or other advanced techniques, relevant to the COVID-19 pandemic, is the following: A person could be thought as having the virus, if that person has travelled to a heavily infected area during the recent weeks. The inference being made here is not a proof that a person has been tested positive for COVID-19, but an indication of the possibility of infection (Skiljic 2021).

The problem of undesired inferences is more evident in IoT due to the increasing amount of data generated and the available data analysis techniques and they constitute a major risk to users' privacy. The subject of privacy protection has been a challenge for researchers since the beginning of the digital age (Foukia et al. 2016). Today, the EU data protection authorities acknowledge the need for the assurance of personal data protection, and in particular the processing of health-related data, which is generally prohibited under GDPR Article 9.[2] As inferences are only predictive and indicative, they may be inaccurate and unverifiable. Nevertheless, they contribute to the creation of

[2] https://gdpr-info.eu/art-9-gdpr/.

user profiles by companies and third parties and could potentially jeopardise people's basic rights and privacy, as the more data that are collected and associated with a user, the more inferences can be made about that user.

2.3 Understanding user awareness and concerns on privacy and IoT

Various studies in the existing literature engage in collecting and analysing the opinions and perceptions of the users of wearable devices regarding the protection of their privacy and the possible risks from the exposure of their personal information without their awareness or consent.

User concerns related to personal data privacy risks are investigated by Lee et al. (2016), where by using a survey with a number of data exposure scenarios in their study they assess user concerns and their results indicate that privacy is at the top of the users' worries when using wearables. On the other hand, the authors have also observed that the users are eager to accept any privacy-related risks, if they consider that the benefit associated with that risk is significant to them. Furthermore, the users' main concerns identified in this study include (a) the disclosure of financial information, which is a user concern related to any possible costs that the user may suffer from the disclosure of stored financial information on their fitness trackers, and (b) location tracking, stalking and physical harm as the result of the use of GPS technology on some wearables. The results of this work provide insights related to how the users of wearable devices discern personal data disclosure. The diversity of our work is that we engage the users of fitness trackers in the process of investigating how their privacy can be compromised from the data created by these devices aiming to make them aware about the various inferences that can be made about them from these data.

The user understanding of the privacy and sensitivity of the data collected by wearable devices is studied in the work of Lehto and Lehto (2017). Using a qualitative research approach to collect data through themed interviews, the study's findings were that overall the participants do not consider the data collected by activity trackers to be private, except in the cases when such data are combined with identifiable information, like name and address. On the other hand, the participants considered health information stored in medical records very sensitive and private. As such, the disclosure of medical information has been identified as a user concern, since users are worried that third parties like banks, insurance companies or employers could potentially benefit from such data when taking decisions regarding loans, insurance rates, hiring new staff, promotions, etc. In our work, we are also interested in the attitudes of wearable devices users and their perception of the privacy levels of the data collected, but at the same time we aim to raise their awareness through a dedicated web application.

The factors taken into account in the privacy calculus of wearable fitness devices are analysed by Cho et al. (2018), who developed a research model based on the privacy calculus theory and used a survey administered to fitness trackers users in order to examine if there is a relationship between the users' intention to disclose personal data and to continue using the wearable device. The results of the survey led to the observation that the users are more likely to continue using the device if the perceived benefits are higher that their privacy concerns identified through the survey. Identified

privacy concerns include the possibility that third parties could gain access to users' personal data, the likelihood that the devices collect too much information about the owners and activity monitoring. What makes our work different is that we support the users to decide whether they wish to continue using a device by educating them about the possible privacy inference risks that stem from their own data and that could not be obvious to the users otherwise.

The users' understanding of the data collection in fitness trackers and their privacy concerns are studied by Fietkiewicz and Ilhan (2020). The authors have used an online survey where current, former or non-users of fitness tracking applications from the EU and USA have participated in order to determine how the different groups comprehend the sensitivity of the data that is collected by these devices and what specific concerns they have in relation to their privacy. The main finding of this study is that users who generally feel insecure about their data privacy online are also more likely to be worried and concerned about the protection of the privacy of their data collected from fitness trackers. User privacy concerns identified through the survey include the likelihood that third parties could gain access to their personal data and that their data could be used against them. While in this work the authors aim to discover the data types with the highest privacy sensitivity that are collected from fitness trackers, in our work we aim to inform the user about any data privacy vulnerabilities that are identified through the dedicated web application.

In their work, Zimmer et al. (2020) employ a survey and semi-structured interviews with current users of fitness trackers in their effort to gain an understanding on the advantages and disadvantages that users perceive from their interaction with these devices. In general, the participants indicate that they have low levels of concerns regarding their privacy and that they consider that the benefits of using a fitness tracker exceed any disadvantages. The outcome of this study according to the researchers is that the users do not perceive data collected from fitness trackers as sensitive, they are not aware of possible threats and they are inclined to share their personal data, like heart rate or step count, as they feel that the privacy risks are low. The diversity of our work is that we exploit fitness trackers user data in order to examine how user privacy can be endangered and make the users aware about the insights that can be extracted about them from these data.

A survey with the goal to investigate the likeness and dissimilarities of fitness trackers users' privacy attitudes from USA and Germany showed that the weight of a number of user privacy concerns varied considerably between the two groups (Ilhan and Fietkiewicz 2020). The introduction of the GDPR in the EU was the driving force for this study, and it has been shown that the European users are using their GDPR rights and have become more responsible of their data. Examples of the identified user privacy concerns include among others the possibility that third parties could gain access to the users' personal data or that their data could be used against them. In our work, we are also interested in the privacy attitudes and concerns of fitness trackers users, but in addition we aim to educate the users about the possible risks and enable them to set their privacy preferences on their fitness trackers accordingly, contributing to the personalisation of the provided services, in respect of their personal data.

An analysis of how fitness tracker users understand the privacy inference risks affiliated with the use of these devices is presented by Velykoivanenko et al. (2021).

Through the use of a longitudinal study, an online survey and interviews with the participants, the authors come to the conclusion that the participants are apprehensive of the types of information that might be inferred about them from their fitness trackers data. The authors go one step further and suggest that one solution to protect the user's privacy is to offer better data minimisation procedures by dropping centralised data collection and by decreasing the granularity of the data collected and sent to the data provider. Contrary to this data-minimisation perspective, we study fitness trackers user privacy from the user viewpoint focusing on increasing the user awareness in relation to the inferences that can be made about them from their data.

The works presented here used methods like surveys and interviews as research tools in order to get insights about how fitness trackers users perceive the privacy risks associated with the data collection and sharing of these devices. A summary of the preceding approaches can be found in Table 1.

2.4 Related work

The use of accelerometer sensors embedded in wearable devices is exploited by Kröger et al. (2019) presenting a number of inferences that are possible from analysing the data collected by such sensors. The identified inferences include activity, behaviour or location tracking. The authors suggest that their findings should be used as a caution to customers and a cause for action to developers and organisations. The possibility of inferences from pedometer sensors that are used to count steps is studied by Yan et al. (2015). The possibility of inferring the user typical routes, for example, going to a coffee shop or a grocery shop, is computed by utilising the steps per minute data from the user's fitness tracker. The Euclidean distance between the steps-tracked sequence and the path query sequence is used to set a threshold value, and as long as this fluctuates, then the user route can be inferred with an accuracy of almost 50%.

The elevation data from fitness trackers are used by Meteriz et al. (2019) to predict the location path of the users, using natural language processing computer vision for the representation of data, and machine learning and deep learning-based techniques to predict and infer personal information, such as frequently visited places. A case study based on fitness trackers is presented by Torre et al. (2016), where a model for inference prevention is built using a Bayesian Network that computes the risk of inference attacks from the combination of known data about users.

A study on the privacy vulnerabilities of fitness trackers is presented by Reicherzer et al. (2017), where machine learning techniques are exploited for the analysis of data from these devices in order to make meaningful inferences about user activities. The results show that it is possible to track users and their activities from their fitness tracker data, creating a threat to their privacy. The possibility of privacy leakages from Bluetooth Low Energy (BLE) communication between fitness trackers and smartphones is examined by Das et al. (2016). As the BLE traffic of fitness trackers seems to be correlated with the intensity of the user activity, the authors show that it becomes possible for a malicious listener to infer the user's activity, by analysing the BLE traffic analysis. They also present their findings regarding the possibility to

Table 1 Summary of approaches on user privacy awareness and concerns in IoT

Citation	Aim and methods	Findings	Privacy concerns
Lee et al. (2016)	A survey to comprehend the users' perception of information disclosure risks in wearables using a list of 72 possible scenarios	Users are concerned about their privacy, but they are prepared to drop their privacy for any benefits	Disclosure of financial information, disclosure of medical information, physical harm due to information leakage, damage to interpersonal relationships
Lehto and Lehto (2017)	A study using interviews with themes sourced from the privacy calculus theory with the aim to discover the privacy concerns of wearable devices users	Users do not consider the data collected by activity trackers as private, but only in cases when these data are combined with identifiable information, like name and address	Location tracking, stalking, third-party access to personal data, third parties can use my data against me, information used for targeted advertising
Cho et al. (2018)	A questionnaire where exploratory factor analysis is applied to investigate if there is a relationship between the intention of fitness tracker users to disclose personal data and to continue using the wearable device	Users are more likely to continue using a fitness tracker if the anticipated benefits are surpassing their privacy concerns	Third-party access to personal data, data used for other purposes than stated, devices collecting too much information, activity monitoring
Fietkiewicz and Ilhan (2020)	An online survey from EU and USA participants who are current, former or non-users of fitness tracking applications to determine their awareness of data collection and their privacy concerns	Users who normally feel insecure about their online data privacy are also more likely to be concerned about the protection of the privacy of their data collected from fitness trackers	Third-party access to personal data, profiling
Zimmer et al. (2020)	A survey and semi-structured interviews with fitness trackers users to determine the pros and cons that users notice from their interaction with their devices	Users have low levels of privacy concerns, they find that the benefits of using a fitness tracker exceed any disadvantages, they do not perceive data collected from fitness trackers as sensitive and they are not aware of possible privacy threats	Third-party access to personal data, third parties can use my data against me

Table 1 continued

Citation	Aim and methods	Findings	Privacy concerns
Ilhan and Fietkiewicz (2020)	A survey on USA and Germany fitness trackers users to investigate the differences between the users' privacy attitudes after the GDPR application in the EU	EU users are more responsible of their data	Third-party access to personal data, third parties can use my data against me, profiling
Velykoivanenko et al. (2021)	A longitudinal study, an online survey and interviews with fitness trackers users to comprehend how they perceive the privacy inference risks of these devices	The participants are apprehensive of the types of information that might be inferred about them	Third parties can use my data against me, damage to interpersonal relationships, inferring various types of information

identify a user by analysing the BLE traffic of her devices, which can depict the unique way a person moves.

The overlooked security and privacy challenges in wearables is the focus of the work by Blasco et al. (2019), where the authors identify a number of inferences that can be extracted from sensors data. According to the authors, fitness trackers become an appealing source of interest for cybercriminals, whose attacks may gain access to users biometric data, enabling identity theft, location information which is a major privacy threat or accelerometer data that can be used to infer user activities. Subsequently, the authors recommend that further research is needed for the consideration of privacy requirements early in the design of fitness trackers and wearables in general.

The limitations of the aforementioned approaches are that even though they show that a number of inferences are possible from fitness trackers data that pose a threat to the users' privacy, none of these works aims to notify the users about them and raise user awareness, and this is what makes our work different from them.

3 Research questions

What we aim to address in this work is to contribute with a tool that will provide awareness to the users about the possible privacy risks and the inferences that can be extracted about them from their fitness trackers data, so that they can set their user privacy preferences in such a way that their personal privacy can be protected adding up to the personalisation of the provided services with reference to their personal data. To accomplish this task, we defined the following research questions.

RQ1 What inferences can be made from the data collected from fitness trackers?

In order to answer this question we use the results from the literature review we performed for this work in combination with our previous research in the area, and we produce a list of possible inferences that pose a threat to user privacy when using fitness trackers. We also aim to find which inferences can be drawn from the data collected from the specific fitness trackers in this study.

RQ2 Are the users aware of the inferences that can be made about them from their fitness tracker data?

For providing an answer to this research question, we conduct an online questionnaire that targets fitness trackers users in order to gain an understanding of: (i) their concerns over their privacy when using their devices, (ii) their awareness of what data are collected by their fitness trackers and how these are being used and shared, (iii) their awareness on the privacy risks form fitness trackers data.

RQ3 Does the PrivacyEnhAction application enhance the awareness of the users regarding the possible inferences that can be obtained from their fitness trackers data?

To answer this question, we provide the same group of fitness trackers users with a number of datasets from the three fitness trackers brands (Fitbit, Garmin and Xiaomi)

Table 2 Indicative list of available commercial fitness trackers

Manufacturer	FT models
Fitbit	Surge, Charge, Ace, Inspire, Luxe
Xiaomi	Mi Smart Band 4C, 5, 6, 7,Redmi Watch 2, Redmi Smart Band Pro
Garmin	Forerunner, Captain, Fenix, Epix, Venu, Vivosmart, Vivofit, Instinct, Quatix
Apple	Apple Watch
Huawei	Band 6, Band 4 Pro, Band 4, Band 4e
Amazfit	Band 7, Band 5, Verge, Nexo, X
Samsung	Galaxy Watch 4
Withings	Scan Watch, Steel HR
Polar	Grit X Pro
Suunto	Peak, Baro

under study. The users are asked to use one dataset for each fitness tracker brand in order to interact with the PrivacyEnhAction app and review the analysis results. Afterwards, they are required to complete an evaluation questionnaire about the app, where they are also expected to answer similar questions to the questionnaire used in RQ2, in order to gain an understanding of whether their awareness regarding inferences has been increased.

4 Privacy policies in fitness trackers

Fitness trackers assist the users in tracking their health, by enabling them to specify what they want to record about themselves, such as their weight, the exercise they perform, the number of steps they take during the day, the distance they walk, how much and when they sleep and their heart rate. This stored information is clear to the users, as these are the data they can see through their profile dashboard. However, further user information is accumulated from the trackers that the user may be unaware of, like the time they wake up, the time they go to bed, their location, timezone, IP address, etc. Even though fitness trackers privacy policies usually state that no data are shared with third parties, this is not always the case as constant user tracking and data collection give fitness tracker companies the opportunity to capitalise on user data with the help of third-party sales (Challa et al. 2017).

A big number of commercial fitness tracker devices are available on the market from different manufacturers, an indicative list of which can be seen in Table 2. For the purposes of this work, we have chosen to employ Fitbit and Garmin fitness trackers after reviewing the available literature, where Fitbit and Garmin devices were identified as the most popular devices (Tedesco et al. 2019). Moreover, Fitbit Surge and Garmin Forerunner appear to have embedded the biggest number of sensors, i.e. PPG, GPS, gyroscope, magnetometer and barometer or altimeter (Henriksen et al. 2018), which means that these devices collect more user data. We have also chosen to include Xiaomi fitness trackers in our study, as Xiaomi appeared in the top five vendors in sales for two consecutive years (2015 and 2016) (Henriksen et al. 2018) and also due to their low cost as our budget was limited.

But what do the privacy policies of the fitness trackers used in this study state regarding data sharing? In this section, we provide a review of how Fitbit, Garmin and Xiaomi fitness trackers address data sharing in their privacy policy.

For the review, we have used the work performed by Perez et al. (2018) where the authors have performed an analysis of the privacy practices that manufacturers provide related to data collection, data ownership, data modification, data security, external data sharing, policy change and policies for specific audiences for six IoT devices and systems, including Fitbit devices. Based on this analysis, we have followed a methodology for gathering the required information about data collection, data sharing, data recipients, privacy policy changes and data handling in case of reorganisation/merge/resale, which are the areas of interest in our research. A summary of the privacy policies review can be seen in Table 3.

4.1 Fitbit privacy policy regarding data sharing

The Fitbit privacy policy states that: "We never sell the personal information of our users. We do not share your personal information except in the limited circumstances described below."[3] The listed circumstances are: (i) when the user agrees to use Fitbit community features like forums, challenges or social tools or directs Fitbit to share her data with third parties, as, for example, when the user gives a third-party application access to her account, or provides access to her employer when choosing to participate in an employee wellness program, (ii) for external processing, to their partners who process user data on Fitbit's behalf in compliance with its policies and (iii) for legal reasons or to prevent harm.

Even though Fitbit's privacy policy states that "we never sell your personal data", it also states later that user data is used for marketing. What this means according to a Fitbit spokesperson is that user data is used only for advertising their own products (McGowan 2021). In the case of a merger, acquisition or sale of assets, the Fitbit privacy policy informs the users that adequate measures will be taken to protect the confidentiality of personal information and give affected users notice before transferring any personal information to a new entity.

According to the Common Sense Privacy Program,[4] a program that evaluates popular applications and services for children aiming to protect child and student privacy, Fitbit fitness trackers do not meet the organisation's recommendations for privacy and security practices. Some of the arguments behind this are, among others, that the trackers collect personally identifiable information (PII), that it is not clear if the data collection or use is bound to the requirements of the device, that the trackers collect geolocation and biometric or health data and also that third parties collect user personal information.

[3] https://www.fitbit.com/global/us/legal/privacy-policy.

[4] https://privacy.commonsense.org/.

Table 3 Comparison of Fitbit, Garmin and Xiaomi fitness trackers privacy policies

Fitness tracker	Data collection	Data sharing	Data recipients	Privacy policy change	Data handling in reorganisation
Fitbit	Account information, location data, usage information, biometrics and fitness info (steps, distance, calories, weight, heart rate, sleep stages, active minutes).	Only shares personal data when user agrees to share, for external processing with their partners according to their policies or for legal reasons and to prevent harm.	Third-party apps, Fitbit corporate affiliates, service providers and other partners.	Fitbit will notify the users before making any changes to the privacy policy and the users will be able to review the revised policy before deciding if they would like to continue to use the services.	Users are informed that adequate measures will be taken to protect the confidentiality of personal information. Affected users will be given notice before transferring data to new entity.
Garmin	Account information, health and fitness information including step count data, heart rate information and sleep data. Activity Data including any activity data like runs, bike rides, swims or other activities recorded with the device.	Only shares data with third-party apps, platforms or service providers with whom the users ask Garmin to share their data with.	Third-party apps, platforms, service providers	Garmin may update their privacy policy from time to time. The users will be provided with notice when this is required by applicable law and will be asked to provide their consent.	Users personal data may be transferred to new entity provided that the new entity will not be permitted to process personal data as per the privacy policy, without providing first notice to the users and obtaining their consent.

Table 3 continued

Fitness tracker	Data collection	Data sharing	Data recipients	Privacy policy change	Data handling in reorganisation
Xiaomi	Exercise information, information recorded by device (activity information, sleep, blood oxygen saturation information and its change, heart rate for each time, resting heart rate, heart rate for whole day, weight), call records for making and receiving calls, your number for sending and receiving SMS, the content of the SMS, the contact name and caller number, MAC address, serial number, firmware version, system time and operating system version of your mobile phone, brand model, information submitted via services, information about near-field communications (NFC) function, other information.	May disclose user personal information to third parties to provide requested products/services. May also disclose personal information to other affiliated companies, to comply with legal obligations, to protect and defend their rights and property or with the user permission.	Xiaomi's ecosystem companies, third-party service providers and business partners, other third parties	If there are material changes to the privacy policy, Xiaomi will notify the users by email or post the changes on their websites or through their software. Users are encouraged to periodically review the website for updates.	Users information may be sold or transferred as permitted by law and/or contract. The users will be notified via email and/or a prominent notice on Xiaomi's website of any changes in ownership, uses of their personal information and choices they may have regarding their personal information.

4.2 Garmin privacy policy regarding data sharing

The Garmin privacy policy includes in the list of possible recipients of the users' personal data various third-party apps, platforms or service providers with whom the users ask Garmin to share their data. In these cases, the third party's handling of the users' personal data is the responsibility of that third party and the users are warned that they should carefully review the third party's privacy policy.

Additionally, Garmin's privacy policy states that: "From time to time, we share or sell activity data in a de-identified and aggregated manner with or to companies that provide Garmin and our customers with content or features for the purpose of enhancing the quality of the content or features they provide and with or to other third parties for research or other purposes".[5] Regarding the possibility of any reorganisation, merger or sale, the Garmin privacy policy clarifies that they may transfer users' personal data to an affiliate, a subsidiary or a third party provided that any such entity will not be permitted to process personal data other than as described in the privacy policy without providing first notice to the users and obtaining their consent.

The Common Sense Privacy Program has only evaluated Garmin Vivofit Jr., and this specific device does not meet the organisation's recommendations for privacy and security practices, for reasons such as the collection of PII, the possibility that user information can be transferred to a third party for advertising and marketing or other purposes.

4.3 Xiaomi privacy policy regarding data sharing

The Mi privacy policy states that: "We do not sell any personal information to third parties. We may sometimes share your personal information with third parties (as described below) in order to provide or improve our services, including offering services based on your requirements. If you no longer wish to allow us sharing this information, please contact us at https://privacy.mi.com/support".[6] The list of third parties includes Xiaomi's ecosystem companies, which are independent entities, other third-party service providers and business partners who may have their own sub-processors, and other third parties with whom Xiaomi may share information in aggregated form. In particular: "To help us provide you with services described in this Privacy Policy, we may, where necessary, share your personal information with our third party service providers and business partners. This includes our delivery service providers, data centers, data storage facilities, customer service providers and marketing service providers and other business partners. These third parties may process your personal information on Xiaomi's behalf or for one or more of the purposes of this Privacy Policy....There may be occasions that third-party service providers have their sub-processors. To provide performance measurement, analysis, and other business services, we may also share information (non-personal information) with third parties in aggregated form". A worrying aspect of the privacy policy is that Xiaomi does not explain what the status of the users' personal information will be in the case

[5] https://www.garmin.com/en-US/privacy/global/.

[6] https://www.mi.com/uk/about/privacy/.

of a merger, acquisition or sale, as the only clarification given is that the users will be notified.

According to the Mozilla Foundation,[7] Xiaomi's Mi Fit Smart Bands do not meet their Minimum Security Standards as they have not responded to how they handle security vulnerabilities. On top of that, Xiaomi has come under fire as it has been secretly collecting personal data from users of its products, and for these reasons, the Mozilla Foundation warns the users against wearing these fitness bands.[8]

5 Possible inferences from fitness trackers data

As most users do not realise the extent of data fitness trackers are collecting, this makes it even more difficult for them to comprehend that these data can reveal more information about them than they can imagine. This section aims to answer the research question: **RQ1** *What inferences can be made from the data collected from fitness trackers?* In order to answer this question, we derived a list of possible inferences that form a threat to user privacy when using fitness trackers, which can be seen in Table 4, based on the literature review we performed.

Activity data: Fitness trackers record the number of steps taken every day by the users, as a measure of their activity level. Activity can be classified using the step index in Table 5 that has been proposed by Tudor and Basset to describe the physical activity in adults based on pedometer readings (Tudor-Locke and Bassett 2004). No or low physical activity is the root behind ill health (Vuori 2004), and therefore, knowledge of this kind of information could be an indication of possible health problems. Information like daily walking step count may potentially reflect people's stable lifestyle and habits or whether someone is at a lower or higher risk of all-cause mortality (Saint-Maurice et al. 2020). Low levels of daily activity could indicate that the user may be suffering from health problems. This information can be used by an interested third party, such as an insurance company, to increase health insurance premiums based on the identified behaviour, for example, that the user does not lead an active or healthy lifestyle.

Activity data can also be used to infer religion. This can be applied particularly for the case of the Orthodox Judaism religion, as on Saturdays believers engage in restful activities to honour the day according to their religion. Even though for most people Saturday is an off-duty and leisure day, if it is observed from the fitness tracker data that the user is usually very active on most days but not on Saturdays, then this could be seen as an indication—not a proof—that the person may be Jewish (Cook 2021). Religion could also be inferred by the time the person wakes up in the morning, since Muslims wake up earlier during Ramadan (Velykoivanenko et al. 2021). Religious or philosophical beliefs are considered as sensitive personal data and could be used in a discriminatory way against a user if obtained by a third party, for example, a potential employer.

[7] https://foundation.mozilla.org/en/.

[8] https://foundation.mozilla.org/es/privacynotincluded/mi-band-5/.

Table 4 Possible fitness trackers inferences

Type of data	Inference	Interested third party	Potential use	Analysis	Sample size	Study
Activity data						
Physical activity	Possible health problems	Insurance companies	Increased rates	T	n/a	Langley (2014)
Amount of physical activity	Chronic diseases	Insurance companies	Increased rates	T	n/a	Booth et al. (2012)
Amount of physical activity	Mortality risk	Insurance companies	Increased rates	T	n/a	Booth et al. (2012)
Amount of physical activity	Human emotions	Employer	Discrimination	T	n/a	Kröger et al. (2019)
Activity, location	Religion	Employer	Discrimination	E	970	Jung et al. (2020)
Activity, location	Religion	Employer	Discrimination	E	227	Velykoivanenko et al. (2021)
VO2max	Fitness level	Insurance companies	Increased rates	E	10	Webster et al. (2021)
Heart rate data						
Resting heart rate	Pregnancy likelihood	Employer	Discrimination	E	8	Aktypi et al. (2017)
Heart rate and respiration	Drugs, cigarette or alcohol use	Insurance companies, employer	Discrimination, increased rates	T	n/a	Peppet (2014)
Heart rate, accelerometer data	Sexual activity	Marketing companies	Targeted advertising	E	227	Velykoivanenko et al. (2021)
Elevated resting heart rate	Health problems, alcohol abuse	Insurance companies, employer	Discrimination, increased rates	E	21853	Cooney et al. (2010)
Elevated resting heart rate	Health problems, alcohol abuse	Insurance companies, employer	Discrimination, increased rates	E	6743	Alhalabi et al. (2017)

Table 4 continued

Type of data	Inference	Interested third party	Potential use	Analysis	Sample size	Study
Low resting heart rate	Bradycardia, medication	Insurance companies, employer	Discrimination, increased rates	T	n/a	Michael Mangrum and DiMarco (2000)
GPS data						
GPS data	Location tracking	Attackers	Targeted home or personal attacks	T	n/a	Bada and von Solms (2021)
GPS data	Frequently visited places	Attackers	Target physical person	E	n/a	Meteriz et al. (2019)
GPS data	Location	Marketing companies	Advertising of products	T	n/a	Fourberg et al. (2021)
GPS data	Health, habits, professional duties	Attackers	User profiling	T	n/a	Cremonini et al. (2013)
GPS data	Location	Many	Political, religious, sexual discrimination	T	n/a	Cremonini et al. (2013)
Sleep data						
Sleep data	Sleep deprecation	Insurance companies, employer	Discrimination, premium rates	T	n/a	Hicks et al. (2019)
Sleep data	Sleep patterns	Marketing companies	Targeted advertising	T	n/a	Valdez (2019)
User weight and height	Obesity	Insurance companies	Increased premium rates	E	n/a	Yao et al. (2020)

Abbr. for analysis: *T* theoretical, *E* experimental. Abbr. for sample study: *n/a* Not applicable

Table 5 Activity levels and steps indices (Tudor-Locke and Bassett 2004)

Activity level	No. of steps
Sedentary	Less than 5000 steps per day
Low active	5000 to 7499 steps per day
Somewhat active	7500 to 9999 steps per day
Active	More than 10,000 steps per day
Highly active	More than 12,500 steps per day

VO2Max data: A number of fitness tracker devices collect the user's VO2Max (cardio fitness level) values. This measurement is thought to be the best indicator of cardiovascular fitness. Monitoring VO2Max over time can assist in establishing whether a person is getting fitter or losing their fitness. Research in the area has shown that low cardio fitness levels are linked with cardiovascular disease, while higher levels are correlated with many health advantages (Fernström et al. 2017; Högström et al. 2016). Therefore, a declining or increasing VO2Max can be used as an indicator of the overall fitness of the user.

Heart rate data: Heart rate data collected by fitness tracker devices are very important and include a treasure of information about our bodies. According to European data protection bodies, heart rate information constitutes part of health data, while under the GDPR, "personal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject".[9] As such, health data including heart rate measurements are considered as a special category of personal data.

Insights about heart rate measurements can assist in observing and understanding one's fitness level, but also to identify possible health problems. The values of resting heart rate, i.e. when the person is sitting and is calm, relaxed and not sick, vary between 60 beats per minute and 100 beats per minute for adults;[10] therefore, a resting heart rate of more than 100 beats per minute is considered high, while a heart rate of less than 60 beats per minute is considered low. A resting heart rate that is below the normal range could be due to a number of reasons. It is a normal situation for a person that is an athlete or a fit and young adult, or it can happen as a side effect of taking a specific medication or from a health condition, such as bradycardia (Jones and Seladi-Schulman 2021; Michael Mangrum and DiMarco 2000). To that end, having a low resting heart rate could indicate that the user is an athlete, she may suffer from bradycardia or is under medication at the time of the readings.

An elevated heart rate could be due to a health condition, exercising at the time of the readings or heavy alcohol consumption, and consequently, the inferences that could be extracted about the user from these data are that the user may be suffering from a heart condition or could be an alcoholic (Cooney et al. 2010; Alhalabi et al.

[9] https://bit.ly/3TPLXgM.

[10] https://www.heart.org/en/health-topics/high-blood-pressure/the-facts-about-high-blood-pressure/all-about-heart-rate-pulse.

2017). The users could face discrimination or increased premium rates, if third parties got hold of such data.

Location data: Location data can reveal individual mobility patterns; when combined with fitness activity information, it may reveal the areas a person mostly works out or even that person's home or work address (Pan 2016). Furthermore, users' fitness activity could reveal their behavioural patterns, including the hours when they are usually away from home. The privacy risk is that if this information falls in the hands of a malevolent third party, then the personal or home safety of the user could be jeopardised. The GDPR acknowledges the location data's unique position as identifiable information by making it part of its definition of "personal data" in Article 4.[11] In the absence of location privacy protection, aggressors can exploit this gap to carry out a variety of attacks. These attacks may include: (i) undesired advertising to users of products near to the user proximity, (ii) physical attacks and harassment or user profiling and tracking, when location data can be used to infer other sensitive information, such as state of health, personal habits or professional duties, (iii) political, religious, sexual persecution and discrimination, in which a person's location is used to restrict his or her freedom (Cremonini et al. 2013), (iv) planned break-in according to the times the user is away from home, (v) stalking.

Sleep data: Sleep tracking is a feature that is supported by most fitness tracker brands, where by using heart rate sensors and accelerometers for movement monitoring, sleep can be detected automatically. Science has long recognised the importance of sleep to physical well-being. People who sleep for less than 6 h have a threefold increased risk of high blood pressure versus those who sleep more, and women who sleep less than 4 h have a twofold increased chance of dying from heart disease than those who sleep longer (Nagai et al. 2010). Moreover, research suggests that a lack of quality sleep is associated with diabetes, obesity and cancer, not to mention worsened mental health and memory. Conversely, sleeping too much is also associated with health problems. Since sleep is fundamental in people's prosperity and physical and mental wellness, lack of sleep and bad quality of sleep have been proven to be linked with health problems, reduced cognitive functioning, bad mood and reduced productivity (Chang et al. 2018).

Furthermore, the extraction of users' sleep patterns from data collected by fitness trackers can be used for user profiling. These user profiles can potentially be exploited by marketing or pharmaceutical companies for targeted advertising, when combined and correlated with other data, like heart rate or interests (Bourreau 2020). A user's personal safety could also be at risk since by tracking sleep patterns, information about when the user ordinarily has the deepest and lightest sleep becomes available, as some fitness trackers collect information about sleep stages. Inferred wake up times may be used by third parties, such as marketing companies, and the user could be targeted for unwanted advertising, since people have better working memory accessibility in the morning close to the time they wake up (Valdez 2019). Additionally, the average percentage of light sleep, deep sleep and REM sleep stages that can be inferred can reveal further insights about user focus capability, mood, memory, use of possible medications like antidepressants, anxiety, depression, etc., while it can be concluded

[11] https://gdpr-info.eu/art-4-gdpr/.

Table 6 Participants demographics

User no.	Gender	Age	FT model and brand
1	Female	45	Fitbit Surge
2	Male	26	Mi Smart Band 4C
3	Female	28	Mi Smart Band 4C
4	Female	45	Mi Smart Band 4C
5	Female	38	Mi Smart Band 4C
6	Female	20	Mi Smart Band 4C
7	Female	38	Garmin Forerunner 630
8	Male	48	Garmin Captain

that people who are sleep deprived are also more likely to make errors and omissions and could then possibly be discriminated against by current or potential employers.

6 Fitness trackers scenarios under study

After the possible inferences that can be extracted from fitness trackers data have been identified, the next step is to find which inferences can be drawn from the data collected from the specific fitness trackers in this study. We also describe the methodology we used in this study in order to collect, examine and analyse the data in the fitness trackers scenarios, following the methodology we proposed in our previous work (Kounoudes et al. 2021) adjusted to suit the current study's needs, which can be applied in other IoT scenarios with minor modifications.

6.1 Data collection process

In this section, we provide details about the data collection process, in relevance to how we gathered our participants and what mechanisms we used for the data collection.

6.1.1 Participant recruitment

We recruited participants by sending email invitations to members of the SEIT Lab[12] of the University of Cyprus that two of the authors are members of. In total, 5 people responded who were fit to participate in the study, meaning that they were over 18 years old and were not diagnosed with any chronic disease. As more participants were required, family and friends of the authors were recruited that fit the criteria. All participants provided their informed consent for submitting their personal data. The details of the participants can be found in Table 6. Before the data collection period started, a meeting was held with the participants in order to inform them about what was required from them, to assist them with setting up the necessary environment by installing the required apps on the their mobile phones and to create personal accounts for the devices.

[12] https://www.cs.ucy.ac.cy/seit/.

Table 7 Fitness tracker datasets downloaded from repositories

Brand	Dataset	Repository
Fitbit	Crowd-sourced Fitbit datasets[a]	Zenodo
Garmin	Run activities[b]	Kaggle
Mi Band	5 years of continuous steps and sleep data[c]	Kaggle
Mi Band	Exported data from Mi Band fitness tracker[d]	Kaggle

[a] https://doi.org/10.5281/zenodo.53894
[b] https://www.kaggle.com/mmaelicke/run-activites
[c] https://www.kaggle.com/damirgadylyaev/more-than-4-years-of-steps-and-sleep-data-mi-band
[d] https://www.kaggle.com/bekbolsky/exported-data-from-xiaomi-mi-band-fitness-tracker

6.1.2 Data collection mechanisms

For the collection of data, we have acquired one Fitbit Surge fitness tracker, five Xiaomi Mi Smart Band 4C devices and two Garmin smartwatches, that were assigned to eight participants, respectively, who were asked to wear them for 24 h a day for a period of 2 months. As more data were necessary for our experiments, we explored various online repositories, such as Zenodo and Kaggle, in order to find additional fitness tracker datasets. Due to the sensitive nature of the data involved, finding suitable public datasets was not an easy task. Still we located a small number of fitness tracker datasets suitable for our experiments, more details of which can be seen in Table 7.

6.2 Data processing and cleaning

In this section, we provide information about how the available datasets were processed and cleaned in order to be ready for the next step of data analysis.

6.2.1 Fitbit datasets

For the first experiment, we employed a Fitbit Surge device owned by one of the participants and we also used the public dataset "Crowd-sourced Fitbit datasets" available at the Zenodo repository (Furberg et al. 2016). This dataset was collected by thirty eligible Fitbit users that participated in an Amazon Mechanical Turk survey, submitting physical activity, heart rate and sleep monitoring data at minute level. In this dataset, different types of data are stored in 18 files in total, where each file contains merged data from the different users. In order to derive suitable data for our experiment in separate sets for each user, we manually processed the dataset by parsing each file by export session ID that corresponds to a unique user. Following this procedure, we acquired a number of user datasets, containing daily physical activity data, heart rate and sleep monitoring data. Each dataset represents a unique user and consists of three files in .csv format. Data processing also required deleting any records containing missing or null values and removing any outliers identified.

6.2.2 Garmin datasets

In this experiment, two volunteers were assigned to wear a Garmin smartwatch for 2 months. Then, each volunteer's data was exported through Garmin Connect using the Request Data Export option. The exported datasets consisted of a number of files in JavaScript Object Notation format (JSON), which were then converted to a CSV format using a JSON to CSV converter tool. Manual examination of the files content assisted in determining which specific data would be useful for data analysis. This process resulted in the acquisition of two files in each dataset at this stage, the first containing general activity data like activity name, activity type, timestamp, duration, distance, calories, startLongitude, startLatitude, avgHr, maxHr, vO2MaxValue, etc., and the second containing sleep data. Again, data processing required deleting any records containing missing or null values and removing any outlier values identified.

6.2.3 Xiaomi datasets

For this experiment, we acquired five Mi Smart Band 4C devices, that were allocated to five participants who wore them for 24 hours for a period of 2 months. When the data collection cycle ended, each participant's data was exported using the Mi Fit account "Export Data" option. The datasets received consisted of a number of folders with data in CSV format, whose content was manually examined in order to evaluate which data would be suitable for analysis. This method led to the inclusion of four files in each dataset, containing activity data, heart rate data, sleep data and user information. Any records with null values or missing data were removed from the files.

6.3 Data analysis techniques

In order to analyse our data, we use statistical analysis and descriptive analytics techniques in our effort to assess and understand the available data. Using the fitness trackers datasets we have at our disposal, we perform Exploratory Data Analysis (EDA), aiming to identify patterns or anomalies on the data using summary statistics and graphical representations, with the intention to identify if any particular data points or the combination of them will facilitate the elicitation of one or more of the designated inferences. EDA is a method that uses data visualisation on datasets in order to determine the relationships of data aiming to find patterns that can reveal hidden information in the data (Rahmany et al. 2020). Correlation analysis, an EDA technique used to measure the strength of the linear relationship between two variables (Sarstedt and Mooi 2019), is applied in order to evaluate the relationships between variables, as any potential connection between variables can enable the extraction of useful information from the data.

6.4 Inference identification in fitness trackers under study

Based on the available data and in line with the analysis performed in the previous section, we undertake the task to identify which inferences can be extracted in accordance

to the inferences list defined in Table 4. It must be noted that the inferences identified in this study are only indications and cannot be used as a verification or evidence. For example, if the available user resting heart rate data can lead to the conclusion that the female user may be pregnant, this inference is not a proof that the particular user is indeed pregnant, but it is only an indication that the user *may be* pregnant.

6.4.1 Fitbit inference detection analysis

Inferences from Fitbit heart rate data: Fitbit heart rate data contain heart rate measurements at 5-second intervals. According to Table 4, using the heart rate measurements we can try to infer: (a) pregnancy possibility, (b) whether the user suffers from health problems in general, (c) alcohol abuse or (d) whether the user is under medication. The procedure described next was adopted for this purpose.

In order to infer pregnancy possibility, information about the user gender is necessary. As this piece of information was not included in the available Fitbit datasets, we did not attempt to extract this insight from the rest of the data, e.g. the resting heart rate.

An elevated or low resting heart rate can assist in extracting inferences (b), (c) and (d). From analysing the available datasets, no information about specific activity and activity times was given, that could be excluded from further analysis. It was then decided to utilise the available sleep data instead. To this extent, heart rate data were combined with sleep data to match sleeping times with corresponding heart rate values and thus extract the resting periods of the user. Using the new combined data, groups of heart rate measurements were created in the cases when there were successive values of above 100 beats per minute and a method was applied to the data to sum up the time between the minimum and the maximum timestamp of each group in order to find the length of time that the elevated heart rate lasted for. From these data, it can be observed that when there are many long periods of time with elevated heart rate, then the inference that can be made is that the user may be suffering from health problems, since the heart rate is elevated during rest time (specifically sleep time). The same procedure was employed for finding the periods of time that the user had a low heart rate (below 60 beats per minute), and if there are many such periods, then it can be inferred that the user may be suffering from bradycardia or may be under medication.

The likelihood of user alcohol abuse can be inferred by using a combination of the available heart rate data and the sleep data, excluding heart rate measurements that fall within the sleeping range. The remaining heart rate data were utilised, creating groups of heart rate measurements when there were successive values above 100 beats per minute and applying a similar method as before for summing up the time between the minimum and the maximum timestamp of each group to find the length of time that the elevated heart rate lasted for. In particular, if the start and end times of these periods follow the same trend, for example, at midnight near the time when the bars close, this could be an indication that the user could be an alcoholic.

Inferences from Fitbit activity data: From the Fitbit daily activity data, we can estimate the activity level of the user. In order to match the activity level of the user to the indices in Table 5, we proceeded by finding the value at which the variable for the Total Steps tended to cluster. Based on this value, we could infer what the activity

level of the user was, and as a result whether the user leads a healthy lifestyle or not. Another inference we worked on using the available total steps data was the religion. Based on this, we calculated the average daily number of steps and we compared them against the average Saturday steps. If the difference between the two values implies that the Saturday activity is unusually low, then we have an indication (not a proof) that this person could be an observant Jew.

Inferences from Fitbit sleep data: Through an accelerometer and the LED located on the back of the watch or fitness device, Fitbit can detect when a user is sleeping and what stage of sleep he or she is in. In order to get insights from the available Fitbit sleep data, we calculated the start and end time of sleep for each calendar day in the sleep dataset. We also aggregated the total sleep time, as well as the total minutes in Light sleep, Deep sleep and REM sleep stages for each day, followed by the estimation of the values at which all these variables tend to cluster. We separated our calculations for weekday and weekend observations, as typically users are likely to have different habits between them. Following this process, we could calculate approximately how many hours of sleep the user gets during the week and the weekend, the time that the user wakes up and goes to sleep and the percentage of his sleep in light, deep and REM stages. Using this information, we can get an insight on whether the user gets enough sleep and her sleep patterns.

6.4.2 Garmin inference detection analysis

Inferences from Garmin activity data: Garmin activity data contain detailed information about user activities, such as running and cycling. Using these data, we were able to extract insights regarding the user's most frequent activities, and then, exploiting the available information about the geographic coordinates (latitude and longitude) of the activity, we applied a reverse geo-coding process in order to find the places that the user's most usual activities take place.

Garmin activity data also contain VO2max measurements, which we exploited over time in order to determine if the specific user has increased or decreased her fitness level. Based on these findings, it can then be inferred whether the user is an athlete, and her overall health status, as the variations of the VO2max values are widely used as an indicator of health.

Inferences from Garmin sleep data: Many Garmin devices have an optical heart rate sensor that utilises an Advanced Sleep Monitoring (ASM) feature, with which users have the ability to track their sleep statistics when wearing the watch while sleeping. Advanced sleep tracking is cut out for recognising when the user falls asleep and wakes up as well as acknowledging the sleep stages taking place throughout the night. Sleep stages include light, deep and REM sleep, which are determined by merging heart rate, heart rate variability, respiration rate, body movement and other measurements.

In our analysis of the available Garmin sleep data, we proceeded by calculating first the total sleep time for each night in the dataset and we determined the regularity of the weekly and weekend sleeping habits of the user. We also aggregated the total minutes in light sleep, deep sleep and REM sleep stages for each night, the total awake minutes of each night, followed by the estimation of the values at which all these variables

tend to cluster. We separated our calculations for weekday and weekend observations, as typically users are likely to have different habits during the week and the weekend.

Following this process, we could infer approximately how many hours of sleep the user gets during the week and the weekend, together with the time the user goes to sleep and the time she wakes up. Similar as before, this information can reveal if the user experiences sleep issues like lack of sleep, and if such information is shared with third parties, such as a current or potential employer, then the user may face unfair dismissal or employment discrimination. Using the inferred data about the average percentage of light sleep, deep sleep and REM sleep stages, one can draw conclusions regarding the user focus ability, her mood or memory, that the user is possibly under medications like antidepressants, that she may be suffering from anxiety or depression, among others.

6.4.3 Mi Fit inference detection analysis

Inferences from Mi Fit activity data: Mi Fit fitness trackers track activities like walking or running, number of steps taken, etc. Using the available Mi Fit activity data, the daily number of steps was exploited in order to estimate the activity level of the user. An analysis on the data was performed and then the value at which the steps variable tends to cluster was determined. Based on this value and the activity indices in Table 5, the activity level of the user could be determined and therefore whether the user leads a healthy lifestyle or not.

The number of daily total steps was exploited in this scenario for the religion inference discussed in Sect. 6.4.1, where we followed the same approach in order to calculate the average daily number of steps and then compared this value against the average number of steps taken on Saturdays. If the difference between the two values implies that the Saturday activity is unusually low, then there is a likelihood that this person could be an observant Jew.

Inferences from Mi Fit heart rate data: The Xiaomi Mi Band collects heart rate measurements at regular intervals set by the user. We followed the same procedure as in the Fitbit heart rate data analysis in Sect. 6.4.1, and we managed to infer whether the user suffers from health problems in general, alcohol abuse and whether the user is under medication. More user information was available in the Xiaomi datasets, including gender details, and therefore, we attempted to use these data to infer pregnancy likelihood. Resting heart rate measurements can be used in combination with the gender to infer pregnancy possibility. Considering that the resting heart rate increases by 30–50% during pregnancy to match the needs of the growing baby (Maganti et al. 2010; Hunter and Robson 1992), we exploited the available personal user information in the Mi Band data that includes the user gender and date of birth, in order to infer the likelihood of pregnancy. We proceeded by combining the available sleep, user and heart rate data, with a view to isolate the data enclosing resting heart rate measurements and upon that we applied a test in order to check if these values fall in the increased by 30–50% range suggesting a possible pregnancy. Information about a person such as pregnancy could reveal information about that person's health and is

classified as special category data in GDPR.[13] To that end, if this type of information is obtained by a third party, it can be used in a discriminatory way against that person.

Inferences from Mi Fit sleep data: The Mi Fit Band uses embedded sensors like accelerometer, gyroscope and PPG (heart rate monitor) to monitor user sleep by tracking body movements and heart rate. The band can also determine whether the user is in light sleep stage, deep sleep stage or REM sleep stage. We followed the same process we applied for analysing the Garmin sleep data, and we manage to infer the hours of sleep the user gets during the week and during the weekend, along with the time the user usually goes to sleep and wake up during the week and the weekend. Likewise the Fitbit and the Garmin scenarios, the information we extracted can disclose whether the user encounters any sleep problems like lack of sleep or insomnia. If such knowledge is shared with third parties, such as the user's current employer or a potential employer, then the user may be confronted with unfair dismissal or employment discrimination. In our analysis, we also calculated the percentage of the user's sleep in light sleep stage, deep sleep stage and REM sleep stage, information that can be used to draw conclusions about the user's ability to focus throughout the day, her memory or mood. This information can also indicate that the user may take medications like antidepressants and that she may be suffering from anxiety or depression.

7 Implementation

As stated in Sects. 1 and 2, the need for tools that will make the users aware of the privacy risks and the possible inferences that can be made about them from their fitness trackers data is now more important than ever, especially under the GDPR requirements.

In our previous work, we introduced PrivacyEnhAction, a web application that aims to inform the users about potential privacy vulnerabilities that emerge from the use of smart water meters and motion sensors (Kounoudes et al. 2021). The front-end of PrivacyEnhAction has been designed following the structure of a web page using html and css styles. The back-end has been constructed using the Flask micro-framework, an open-source framework supporting the development of Python-based web applications. In this system, we have collected and trained approximately 606,000 smart home motion sensor records and 3100 smart water meter records with a number of unsupervised machine learning algorithms to build a model for the extraction of inferences from these types of data. All code was written in Python and implemented in the Spyder IDE environment.

We have now extended this tool by adding the three fitness trackers to the list of smart devices whose data can be analysed.

Focus group and data collection. In order to enhance the functionality of the PrivacyEnhAction application for the needs of the fitness trackers case, we formed a focus group with 5 researchers from the University of Cyprus, who also agreed to participate in the data collection process by providing their fitness tracker data to assist the

[13] https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/what-is-special-category-data.

research work. We performed structured, open and group interviews with the focus group as a whole and with each member individually in order to record their concerns in relation to the protection of their privacy and to gather requirements for the development of the new system functionality. The extensions that have been made to the systems are described next.

7.1 Extensions to the PrivacyEnhAction tool

The tool has now been extended to include Fitbit, Garmin and Mi Fit 4C fitness trackers in the list of the available devices. The additional implemented functionalities consist of the following:

1. Inference detection analysis for Fitbit Surge
2. Inference detection analysis for Garmin Captain and Garmin Forerunner 630 models
3. Inference detection analysis for Xiaomi Mi Smart Band 4C

The code written in Python has been changed to include the new changes, where, depending on the selected device, the corresponding modules are called to process the files that are uploaded by the user. Using a number of statistical analysis methods, the application displays to the user the data-driven conclusions and possible inferences that can be drawn from her data. Then, the user can select to view more information about each inference type, along with the possible risks that exist in relation to their privacy. Dedicated templates have been developed for each option that are rendered accordingly. The user interface has been adapted to reflect the new additions to the system following Nielsen and Molich's 10 user interface design guidelines (Molich and Nielsen 1990) retaining all graphic representations and text across every system template.

Users of these fitness trackers models can upload their data to the application through the interface, after they have exported them from their corresponding account dashboard, in order to analyse the data and view the possible inferences that could be extracted about them and be informed about the potential privacy risks that these inferences entail. Figure 1 illustrates in a screenshot the Inference Detection Analysis page of the application, where the users can select the device they want to test for inferences.

When the data are processed, the application presents to the users the different types of inferences that could be drawn from their data, as illustrated in Fig. 2, in which case the user has analysed Fitbit data. The privacy risks for each inference type are demonstrated by clicking on the corresponding button through the use of textual information and graphs related to the user's data, as well as further educational information, messages and links, as portrayed in Fig. 3.

In the illustrated example regarding the inferences that can be extracted from the user's Fitbit heart rate data, the user is informed about what information can be revealed from the heart rate in the first part of the interface. Further down, the number of days and records in the processed dataset are displayed. In the next block, the user can view the number of incidents where her heart rate was below 60 bpm or over 100 bpm (low and high heart rate inferences, respectively) during the days processed, as well as the

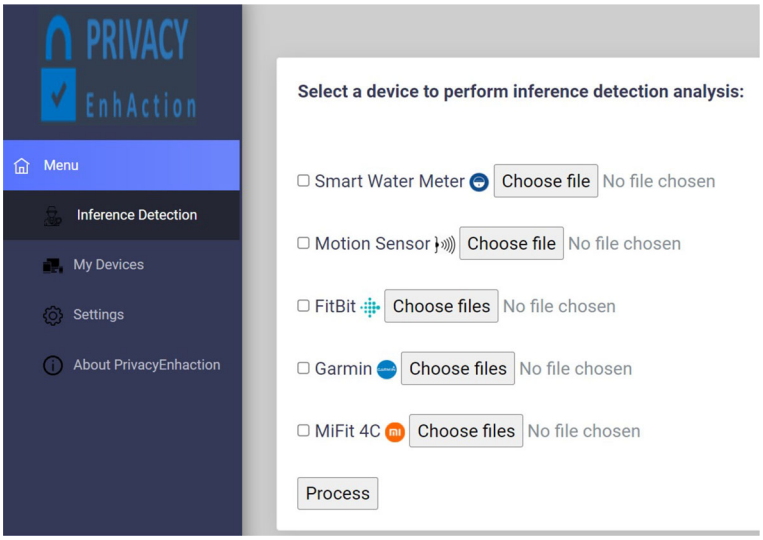

Fig. 1 Screenshot from the PrivacyEnhAction inference detection analysis page

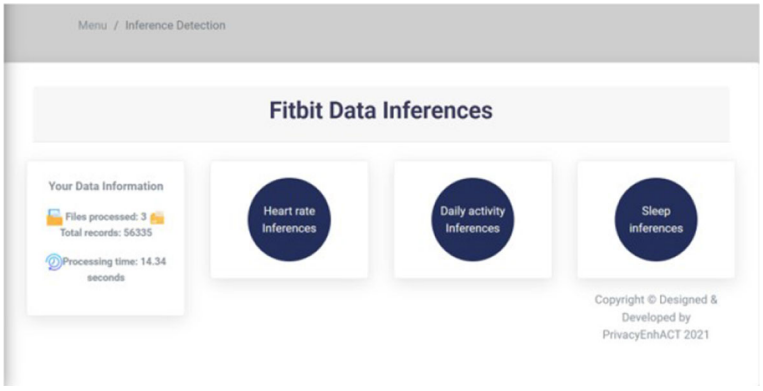

Fig. 2 Screenshot from PrivacyEnhAction Fitbit data inferences results page

total time that these events lasted for. The graphs of low heart rate and high heart rate over time for this time period are then presented. The user is then informed about the privacy risks and the insights that could be drawn about them from their heart rate data and by clicking on the blue sign we aim to increase the user's awareness by letting the user know how this information could be used by interested third parties.

Inferences that could be obtained from Garmin and Mi Smart Band 4C fitness trackers are presented to the users in a similar manner.

Fig. 3 PrivacyEnhAction: Fitbit inferences from heart rate data

8 User evaluation

In this section, we describe the experimental process we performed in order to evaluate the impact of PrivacyEnhAction to the awareness of the users in relation to the privacy risks and the inferences that could be drawn about them from their fitness trackers data.

8.1 Material and methods: empirical approach

We followed a three-step empirical approach that is described below:

1. **Step 1**: A first questionnaire whose aim is to collect information about the awareness and the concerns of fitness trackers users regarding their privacy when using fitness trackers was created and distributed (*Questionnaire on fitness trackers user privacy concerns* (Kounoudes 2022a)).
2. **Step 2**: The participants are provided with the datasets collected during the data collection process described in Sect. 6.1, after being anonymised, and are requested to use them in order to interact with the PrivacyEnhAction application. The existing datasets were used for evaluation purposes and in order to let participants use the application without providing their own personal data.

3. **Step 3**: A second questionnaire was created and distributed to the same group of users as in Step 1 and Step 2. By using a number of questions similar to the ones in the first questionnaire, it aims to assess if the users' awareness and privacy-related concerns have changed (i.e. improved) after interacting with the application (*PrivacyEnhaction Evaluation Questionnaire*(Kounoudes 2022b)).

It has to be noted that participants had to complete all steps in order for their response to be considered as valid. The analysis of the two questionnaires results aims to address research questions **RQ2**: *Are the users aware of the inferences that can be made about them from their fitness tracker data?* and **RQ3**: *Does the PrivacyEnhAction application enhance the awareness of the users regarding the possible inferences that can be obtained from their fitness tracker data?* For the analysis of the results, we used IBM SPSS Statistics for the generation of data descriptive statistics and item-level results of each question.

8.2 Research participants recruiting

The User Evaluation survey was distributed through email communication in order to recruit participants. No monetary or other incentive was provided as a reward for answering the survey. The email provided information about the research goals, stating the objectives of the study and it also included the links to the survey questionnaires, the PrivacyEnhAction application and the share link of the available datasets and the application user guide. No screening criteria were applied, other than that the participants were owners of fitness trackers or smartwatches. A total of 47 responses were collected. Out of these responses, 17 participants did not complete the second questionnaire and as such these data were removed. Finally we had 30 valid responses which were used in our analysis.

8.3 Analysis of responses: user privacy concerns

In our initial questionnaire, the first section consists of social and demographic questions, like gender, age, education level and profession. We used the gender as a demographic variable in order to determine if there exist any opposing views in the attitude and awareness of the privacy risks of the use of fitness trackers between male and female users of the study. In the literature, age is considered as a negative factor in the acceptance of technology (Peek et al. 2014), and for that reason we also used this as a demographic variable in order to find out if it can affect the results in relation to the awareness of the users of the privacy risks of the use of fitness trackers. The second section includes questions regarding information about fitness tracker ownership, such as frequency of using a fitness tracker, length of time of ownership of a fitness tracker and the fitness tracker brand being used. The third part consists of questions related to the user's attitudes against reading the fitness tracker's privacy policies and changing the default privacy settings. The fourth section includes questions about the user awareness on fitness tracker data collection and sharing, while the fifth section consists of questions related to the user's awareness on the privacy risks form fitness trackers data. The sixth part of the survey includes questions about the user's privacy

Table 8 Information about fitness tracker brands being used by the survey participants

Brand	Frequency	Percent
Apple	2	6.7
Fitbit	4	13.3
Garmin	6	20
Samsung	5	16.7
Xiaomi	5	16.7
Other	8	26.7

Table 9 Information about length of time of using a fitness tracker

Answer	Frequency	Percent
The past 3 months	6	20
The past 6 months	2	6.7
The past year	4	13.3
The last 2 years	4	13.3
More than 2 years	14	46.7

concerns when using fitness trackers. The next section contains questions regarding the users attitudes in relation to good uses of data if shared, and the last section gathers the user opinions about the importance of the creation of tools that would make the users aware of how their data are collected and shared by smart devices.

8.3.1 Demographics and other results

In the data analysis, the gender breakdown achieved was 66.7% male and 33.3% female. The majority of the participants are employed at the Engineering and Manufacturing sector (30%) and the IT sector (26.7%), followed by the education (10%), accountancy, banking and finance (6.7%), business, consulting and management (6.7%), environment and agriculture (3.3%), healthcare (3.3%) and other sectors (13.6%).

In relevance to the fitness tracker or smartwatch brand being used, Table 8 shows the frequency and percentage of participants using each fitness tracker brand. The length of time that the participants have been using their fitness trackers or smartwatches is reported in Table 9.

The analysis of the responses in the third section of the questionnaire shows that 80% of the participants in our questionnaire does not read the privacy policy of their fitness tracker, 86.6% does not read the terms and conditions and 70% has never changed the default privacy settings (Fig. 4).

The aim of the next section of the questionnaire was to examine the participants' perceived knowledge and awareness of the data collection process performed by fitness trackers or smartwatches, as well as to see if they acknowledge the types of data collected and what happens to that data afterwards, using a 'Yes', 'No' and 'Maybe/I don't know' type of question. The results in Fig. 5 show that a big percentage of the participants (83.3%) is aware that personal data are collected by fitness trackers, but

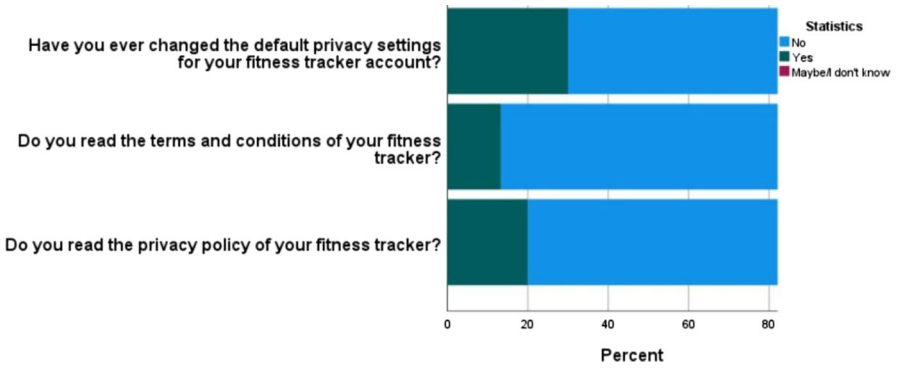

Fig. 4 Users attitudes with regard to privacy policies

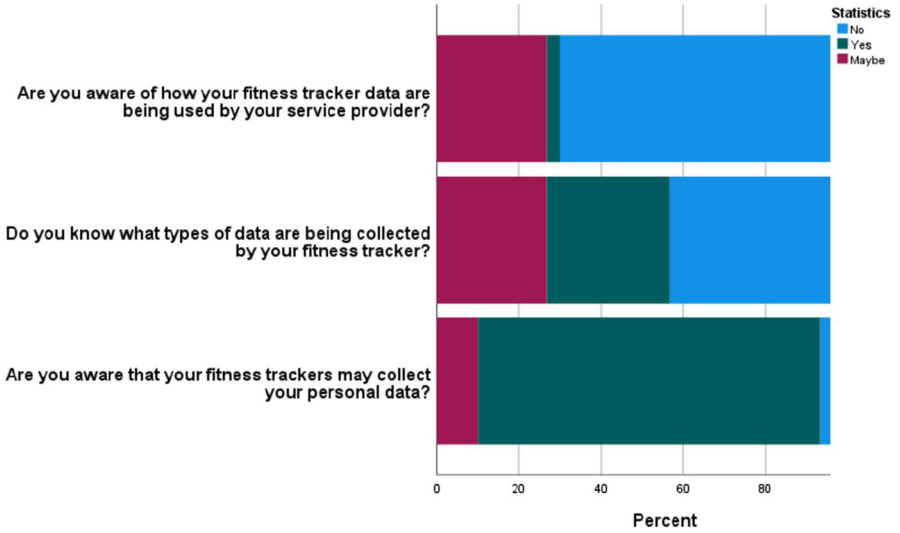

Fig. 5 User awareness on fitness tracker data collection and sharing

only a 3.3% understands how these data are being used by the service provider and a 30% of the participants is aware of the types of data that are being collected by fitness trackers.

8.3.2 User awareness on privacy risks

In this section, in order to understand the users' awareness and perception of the possible privacy risks emerging from the use of fitness trackers, the participants were presented with a number of events and were asked to give their opinion regarding the possibility that they could possibly occur, using a 5-point Likert scale with values ranging from 1 = Very unlikely to happen to 5 = Very likely to happen.

The test of normality showed that our data are normally distributed with $p = 0.38$. The overall mean score of the Likert scale in Sect. 5 that consists of 16 items is 3.37,

Tests of Normality

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
S5	.127	30	.200	.964	30	.380

Descriptive Statistics - Q1

	N	Minimum	Maximum	Mean	Std. Deviation
S5 Over all Mean	30	1.81	4.88	**3.3771**	.77898
Valid N (listwise)	30				

Fig. 6 Normality test and descriptive statistics in Sect. 5 data

which translates to the average response of the users in relation to their awareness about the possible inferences in relevance to the scenarios they were presented with as being "Undecided" (Fig. 6). As can be seen in that figure, it seems that the participants are aware of a small number of inferences that can be drawn from their fitness tracker data. For example, in regard to the scenario "Marketing companies can use fitness tracker data in order to send you specific advertisements regarding running shoes", 68.2% of the users reported this as "Very likely to happen" and 18.2% as "Likely to Happen", while none of the respondents responded with "Very unlikely to happen" or "Unlikely to happen". This is quite predicted as online targeted advertising has shown great market potential (Yan et al. 2009) and is widely used today. In another case, the scenario "A murder can be solved by using the victim's fitness tracker data, such as heart rate data" has been acknowledged as "Very likely to happen" by 54.5% and as "Likely to Happen" by 27.3% by the participants (none of the participants responded with "Very unlikely to happen" or "Unlikely to happen"). This is explainable as in the recent past there have been many murder cases reported in the news where the data from the fitness tracker worn by the victim have assisted in the determination of the exact time of death and led to the murder being solved (Hantke and Dewald 2020; Lovejoy 2021).

The participants' opinions diverged regarding religion inferences, as 22.7% have responded to this scenario as "Very unlikely to happen" and 13.6% as "Unlikely to happen", but a 40.9% is undecided about this possibility. Similar levels of responses across all answers were observed for the scenario "Your fitness tracker data can be used to make the assumption that you are an alcoholic", where the answers were spread with 22.7% for "Very likely to happen" and "Likely to Happen", 18.2% for "Undecided", 13.6% for 'Unlikely to happen' and 18.2% for "Very unlikely to happen".

In relation to the effect that the participants' gender has to answers, further analysis on this section's questions has shown that the gender is not correlated with the user awareness about the possible inferences that can be extracted from fitness trackers data. Furthermore, using the ANOVA test, we investigated the effect that age has on the responses, and we deduced that age has a significant impact on the following statements:

1. *Insurance companies can increase the premium rates of clients based on their low activity levels from their fitness tracker data* ($F = 3.335$, $p = 0.026$): For this scenario, younger participants (aged 18–25) have responded with a mean score of 4.67, thus showing that they believe that such a scenario is very likely to happen, while older participants (aged 56–65) have responded to this question with a score of 1, i.e. as very unlikely to happen and participants aged between 46 and 55 have responded with unlikely to happen.
2. *Marketing companies can use fitness tracker data in order to send you specific advertisements regarding running shoes* ($F = 5.477$, $p = 0.003$): In this scenario

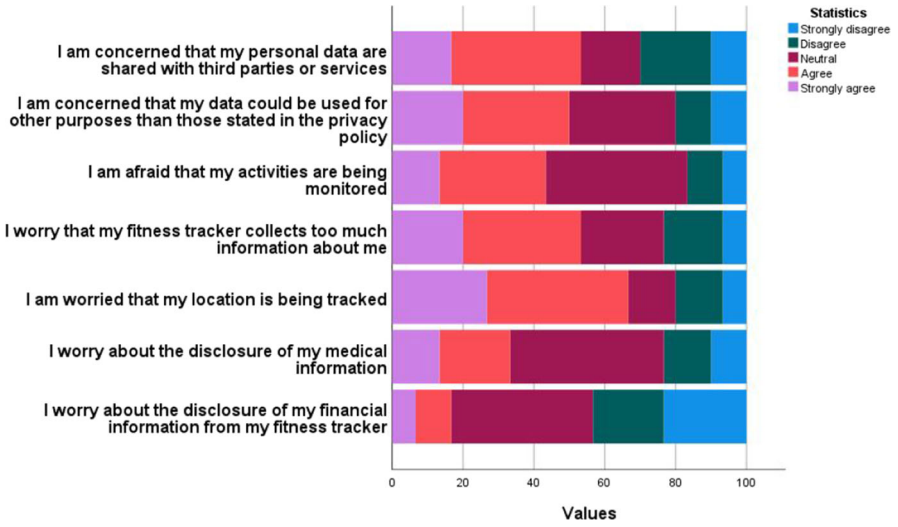

Fig. 7 Examples of user privacy concerns regarding the use of fitness trackers

younger participants believe that it is very likely to happen, while older participants are more reluctant to accept it.

3. *Marketing companies can use fitness tracker data in order to send you specific advertisements regarding coffee brands* ($F = 2.941$, $p = 0.04$): Older participants believe that this scenario is likely to happen while younger participants are more sceptical.

8.3.3 User privacy concerns

In order to understand the privacy concerns of the participants, they were asked a number of questions about specific concerns related to the use of fitness trackers, using a 5-point Likert scale with values ranging from 1 = Strongly disagree to 5 = Strongly agree. The concern that worries the participants the most is the possibility that their personal information may be used for target advertising, where 33.3% of the participants strongly agree and 23.3% agree with the statement, followed by the fear that their location in being tracked, with 26.7% of the participants responding with Strongly agree and 40% with Agree (Fig. 7). Further analysis on the questions in this section shows that gender does have an effect on the users' privacy concern.

In relation to the participants' awareness to the data collected by fitness trackers, location is the most popular answer to this open question (60%), followed by heart rate (40%) activity type (30%) and health data (30%) (Fig. 8).

8.4 Analysis of questionnaire on PrivacyEnhAction application evaluation

In the second questionnaire, the participants had to answer the same set of questions regarding their awareness on the privacy risks and the possible inferences that could be

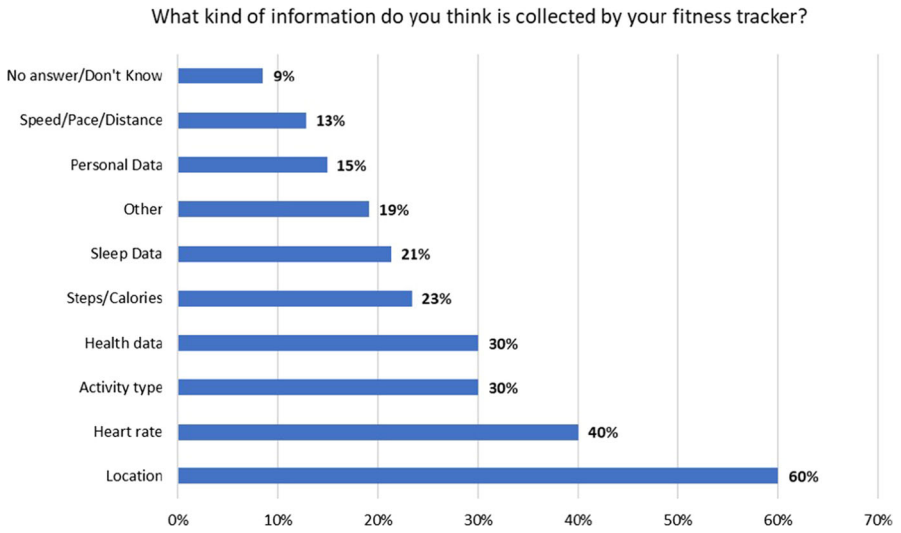

Fig. 8 Users awareness regarding the data collected by fitness trackers

Table 10 Overall mean scores on user awareness before and after interacting with the PrivacyEnhAction application

	Min	Max	Mean	SD
Before	1.81	4.88	3.33771	.77898
After	3.00	5.00	4.0417	.50704

extracted from fitness trackers data, as in the first questionnaire, in our effort to seek an answer to Research Question **RQ3**: *Does the PrivacyEnhAction application enhance the awareness of the users regarding the possible inferences that can be obtained from their fitness tracker data?* In Sect. 8.3.2, we showed that the overall mean score of the same section in the first questionnaire is 3.38. The overall mean score of the same set of questions in the second questionnaire, i.e. after the participants have interacted with the PrivacyEnhAction application, is 4.04, as can be seen in Table 10.

To verify our results, we conducted a paired sample *t* test in order to compare the degree of the users' privacy awareness before and after interacting with the PrivacyEnhAction application, using the same set of questions that exist in both questionnaires regarding user awareness on privacy risks. Using Cronbach's alpha indicator, we evaluated the reliability of the two Likert scale set of questions of the questionnaires. The results for questionnaire 1 demonstrated good internal consistency with a score of 0.876, while the results for the set of questions of questionnaire 2 showed acceptable internal consistency with a score of 0.768. The results of the paired sample *t* test suggest that there is a statistically significant difference between the level of the users' awareness before and after their interaction with the PrivacyEnhAction application, as shown in Table 11. A p value below 0.05 was considered statistically significant. The pairs of questions that differ are the following:

Table 11 Paired sample *t* test results

	Mean	SEM	*t*	*df*	Sig
Pair 1	− 1.7	0.3	− 5.667	29	0.001
Pair 2	− 0.967	0.309	− 3.13	29	0.002
Pair 3	− 0.5	0.302	− 1.654	29	0.054
Pair 4	− 0.767	0.261	− 2.935	29	0.003
Pair 5	− 0.433	0.27	− 1.606	29	0.06
Pair 6	− 0.433	0.223	− 1.941	29	0.031
Pair 7	− 0.167	0.369	− 0.452	29	0.327
Pair 8	− 0.433	0.345	− 1.257	29	0.109
Pair 9	− 1.733	0.332	− 5.222	29	0.001
Pair 10	0.133	0.243	0.548	29	0.294
Pair 11	− 0.867	0.321	− 2.703	29	0.006
Pair 12	0	0.275	0	29	0.5
Pair 13	− 2	0.303	− 6.595	29	0.001
Pair 14	− 0.633	0.305	− 2.076	29	0.023
Pair 15	− 2.267	0.325	− 6.975	29	0.001

- Owners of fitness trackers can be discriminated against due to their religion or race rooted in assumptions extracted from their fitness tracker data.
- Insurance companies can increase the premium rates of clients based on their low activity levels from their fitness tracker data.
- The exact fitness activity movements of a fitness tracker user can be tracked from fitness tracker data.
- Marketing companies can use fitness tracker data in order to send you specific advertisements regarding running shoes.
- Assumptions about your religion can be made from your fitness tracker data.
- Your fitness tracker data can be used to make the assumption that you are an alcoholic.
- Your fitness tracker data can be used to make the assumption that you suffer from short-sightedness.
- Your fitness tracker data can be used to make the assumption that you suffer from heart problems.
- Your fitness tracker data can be used to make the assumption that you suffer from insomnia.

We further analyse if the users will take specific actions after their interaction with the application in relation with the use of their fitness trackers. In particular, 53.3% of the participants said that it is very likely that they will change the default privacy settings of their tracker, while 23.3% responded with Likely. Regarding the statement "Allow the tracker provider to use your data for specific purposes that you choose", 56.7% and 23.3% of the participants responded that this is Very Likely and Likely, respectively. If we compare the participants' answers in percentages in Fig. 9 with their responses in Sect. 8.3.3 in relation to the participants' attitudes against fitness trackers privacy policies, terms and conditions, etc., we can see that PrivacyEnhAction

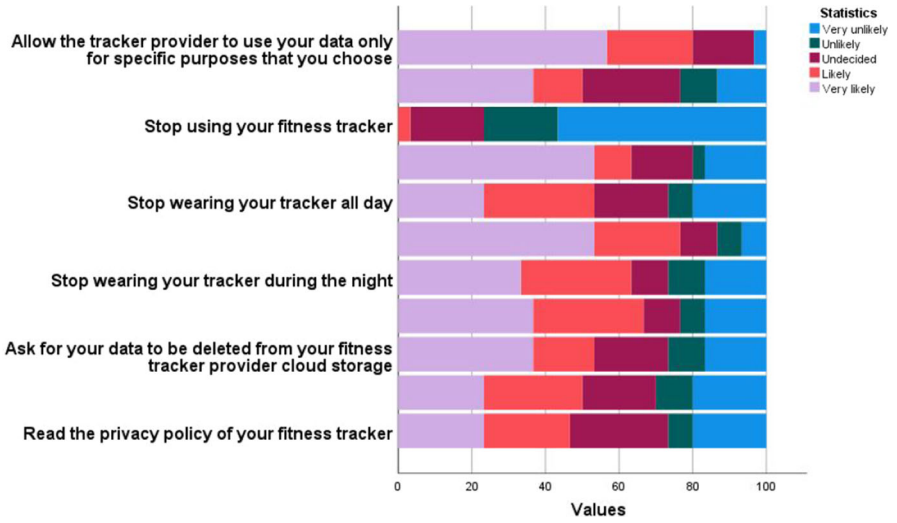

Fig. 9 Examples of users possible actions after using PrivacyEnhAction

has increased their awareness, as 26.6% more of the participants will now read the privacy policy of the trackers, 36.7% more will now read the terms and conditions and 46.6% more will now change the default privacy settings of their account.

In the next section, the participants had to provide their feedback with regard to their interaction with the PrivacyEnhAction application. According to the responses, 83.3% of the participants think that their awareness regarding the use of their personal data from their fitness trackers has been increased after they have used the app. Furthermore, 56.7% of the respondents find that their awareness about the possible inferences that can be made about them and their habits from their fitness trackers data has been increased to a high degree, while 30% think that it has very much been increased. As to the users' privacy concerns, 86.6% think that the use of the application has increased their awareness ranging from very to a high degree, while 10% think that it has not increased their awareness at all. It is, however, very important to mention that all the participants have reported that they believe that PrivacyEnhAction is a useful tool for informing them about the possible inferences that can be extracted about them from their data that may violate their privacy and to provide user awareness.

9 Discussion and limitations

In this work, our research was guided by the ambition to create a tool that will increase the users' awareness in the area of fitness trackers with reference to what information can be figured out about them from the data created and shared by their fitness trackers or smartwatches. Our intention was to educate the users about the possible risks and enable them to set their privacy preferences on their fitness trackers accordingly, contributing to the personalisation of the provided services, in respect of their personal data.

In order to reinforce our research, we have performed a review of how each fitness tracker brand used in our study addresses data collection and sharing, and how these are presented in the privacy policy. Even though privacy policies should assist the users to make informed decisions regarding the use of their device, current policies lack usability, as users tend to ignore them and thus miss important information which includes details about providing their consent (Reinhardt et al. 2021).

In regard to data collection, Fitbit and Garmin collect account, health, fitness, geolocation and device information like number of steps, distance travelled, calories burned, weight, heart rate, sleep stages, active minutes, as well as additional information that the users choose to provide. These types of information have been exploited in this study in order to increase the user awareness about the inferences that may be extracted about them from the data collected from their devices. In the case of Xiaomi, the information collected far exceeds the necessary information for the service a fitness tracker is supposed to offer, as the devices also collect the MAC address, serial number, firmware version, system time and operating system version of the mobile phone connected with the Xiaomi Wear App, as well as information about SMS or message reminder functions, call records for making and receiving calls, the number of the mobile phone in use, the content of the SMS, the contact name and caller number. These types of information were not analysed in this study at this stage; therefore, as future work, we are investigating the privacy vulnerabilities of these information types in order to raise user awareness, as it is very important that the users become aware of what information is being collected from their fitness tracker, considering that a big amount of personal information is at risk.

Data sharing is the next aspect that users should be vigilant for, as, for example, by granting access to a third-party app to their account, then the use of the account information will be governed by the third party's privacy policy, and not the fitness tracker's. It is crucial for the users to be aware about this term written in the fitness trackers policies, and it raises the importance for application providers to disclose their privacy policies in a clear and easy to read manner, enabling the users to protect their privacy (Kang and Jung 2021).

In order to address **RQ1**, the literature review we performed in the area assisted us in the formulation of a list of possible inferences that pose a threat to user privacy when using fitness trackers. We limited the inference list to those inferences that we could identify at the time that the research was taking place based on the available data we had. Using the list, we implemented the new functionalities for the PrivacyEnhAction application for the three fitness trackers we had at our disposal, and the results showed that multiple data points can be used to infer and possibly predict health, fitness status, pregnancy, religion, etc. Not surprisingly, Prince in her work (Prince 2021) explains very effectively that a big amount of health information can be inferred from location data.

The findings of our study demonstrate that the use of tools, like PrivacyEnhAction, can assist in the increase in the users' privacy awareness when using smart devices. In our work, we aimed to gain an understanding of fitness trackers users' awareness and concerns regarding their privacy when using fitness trackers, through the first questionnaire. The results have shown that even though a big percentage of the users are aware that their trackers may collect their personal data, they do not take any

action to minimise any possible risks, such as by altering their fitness trackers privacy preferences or by reading the privacy policies of their trackers in order to get informed. This finding agrees with prior research in the area, where results show that fitness tracker users do not change the default settings of their devices and they do not read their privacy policies (Velykoivanenko et al. 2021; Gabriele and Chiasson 2020); even though the majority of the respondents agrees with the privacy policies and terms of service, they continue to skip them due to information overload (Sigmund 2021), and also because they consider them to be annoying and lengthy (Obar and Oeldorf-Hirsch 2020). This observation also indicates that personal data privacy awareness is not equivalent to the understanding of personal data privacy protection (Chen et al. 2013). We also found that only a small portion of the sample understands how the personal data collected by fitness trackers are being used by the service providers. This is in line with the work of Vitak et al. (2018), which showed similar results from a survey of Fitbit and Jawbone users about the user privacy concerns in relation to tracking and sharing.

Our participants responses in relation to their awareness about the inferences that could be extracted from their data and how these could be used by third parties showed that the users are apprehensive only for a few of the scenarios that they were presented with, while overall they seem uncertain about the possibility of the extraction of the presented inferences. A previous study in the area by Velykoivanenko et al. (2021) has linked the participants beliefs with their understanding of the embedded sensors in their device and the data collected by those sensors. This could justify the participants responses in relation to the scenarios presented to them and enable us to give an answer to research question **RQ2**: *Are the users aware of the inferences that can be made about them from their fitness tracker data*, where we can say that the user awareness depends on the scenario, but in general the users are not aware of the possible inferences that could be extracted about them.

The results of the analysis of the second questionnaire have produced more comprehensive conclusions as to whether the users' interaction with the PrivacyEnhAction application has increased the users' awareness (**RQ3**). In regard to the inferences that could be extracted from the users' data or how these data could be used by third parties, it has been observed that the participants seem to be more educated and more aware about them after interacting with the application, as the mean value of the responses in the relevant section of the questionnaire is "Likely to happen". Comparing this with the mean value "Undecided" in the same set of questions in the first questionnaire, we can safely conclude that the users' interaction with the PrivacyEnhAction application has increased their awareness regarding the inferences that could be extracted from their data and how these data could be used by third parties. This demonstrates that the privacy education that PrivacyEnhAction intends to bring to the users through its graphical interfaces, the pop up messages and the educational tips it provides, seems to be working and proves that embedding privacy education in an application with simple and clear descriptions is a required feature for enhancing user privacy awareness and education (Velykoivanenko et al. 2021; Aktypi et al. 2017). As users appear to be ignorant of how their personal data could potentially be used, it is important that education mechanisms take the context into consideration when including the user in

the process. For the fitness trackers example under study, this is essential due to the sensitive types of data collected.

Our results showed a positive relationship between the use of a privacy awareness mechanism and the increase in the awareness of the user about the possible privacy risks of using a fitness tracker. Enhancing the users' control over their privacy by assisting them to understand the data practices of the smart devices they own, adds to the strengthening and boost of their privacy awareness. These findings are aligned with earlier studies where it is reported that privacy awareness mechanisms like data dashboards, similar to PrivacyEnhAction, are well perceived by users in terms of effectiveness and easiness to use, and also due to the detailed information provided (Thakkar et al. 2022). The communication of the potential privacy risks to the users and its effect to the users' awareness is also investigated in our study. The results showed that the users' privacy awareness had a positive relationship with informing the users about any potential privacy risks, being in line with previous studies which give directions for the creation of privacy awareness mechanisms (Vemou et al. 2014).

The findings from this study provide valuable insights for the users of fitness trackers in our effort to increase their awareness; however, despite the possible privacy risks, the inferences that can be extracted from fitness trackers data can also have a positive impact to the users. Tracking the daily activities of a user can help to enhance the user's health in the long term as the user can be assisted to reach her fitness goals (Wu et al. 2016). The observation of personal health data collected from fitness trackers can lead to the detection and prevention of diseases, such as COVID-19 (Gross et al. 2021), heart diseases (Kaiser et al. 2016; Al-Makhadmeh and Tolba 2019) or diabetes (dia 2017), and even sleep problems (Sathyanarayana et al. 2016). In all cases, it is important that the users understand the privacy complications of using fitness trackers and the potential inferences from personal data, while at the same time balancing the benefits of their functionalities.

Limitations. We acknowledge that this research may have some limitations; however, it could provide the means for further research in the relevant area. First, the size of the participants sample cannot represent the smart devices user population, even though we tried to recruit a diverse sample of participants in terms of demographic variables in order to increase the probability that the results we are aiming for have been indicated by at least one of our participants. Hence, the statistical analysis performed on our sample provides only indications; it is, however, useful in analysing our results. Even though our participant recruitment methods were designed to minimise response bias, by electronic mails to random and known addresses at public and private universities at Cyprus and abroad, the sample is considerably more educated than the general population. This parameter may bring bias to the results in terms of the knowledge and the awareness of the users regarding the privacy risks.

Another limitation is the reluctance of a portion of users to be educated about the privacy risks of using a fitness tracker, as they consider that the benefits of their devices are more important than any possible risks and are therefore uninterested in anything other than the provided services. When starting our research, we acknowledged that these types of users will probably not going to use the PrivacyEnhAction application. In order for the users to seek technologies or applications that educate them about fitness trackers privacy risks, policy makers and regulatory organisations should engage in

actions aiming to increase the privacy awareness of users of smart devices in general. To that end, it is essential to provide tools and methods that enable the increase in privacy awareness.

10 Conclusions

In this work, we have investigated the possibility of getting insights and extracting inferences about the users from their data collected from fitness trackers. We present a list of possible inferences that pose a threat to user privacy through the use of fitness trackers and we utilise our privacy tool, PrivacyEnhAction, as a means for increasing the users' awareness about the privacy risks that emerge from the data collected by their devices, in order to enable the users to set their privacy preferences in an appropriate way, contributing to the personalisation of the provided services in connection with their personal data, while protecting their privacy at the same time. The results of our experimental research have showed that the interaction with the PrivacyEnhAction application can increase the user awareness on the possible inferences that can be obtained from their fitness trackers data.

The methodology used can be adapted in other scenarios as well as it is not bound to smart home or fitness trackers scenarios. We believe that the results of our experimental research can act as a stepping stone in a common effort to bring the smart devices owners in the heart of the privacy risks awareness process with the aim to increase their knowledge and guide their attention towards those actions that can protect them from potential harm, and also for the provision of better services to the users. For future work, we plan to automate the step of data export without the user having to download the data from her dashboard and then upload it to the application. This will be done though the corresponding fitness tracker API and will reduce the number of steps the user has to make, making the process faster and more reliable.

Author Contributions The authors did not receive support from any organisation for the submitted work.

Data Availability The results of the questionnaires and the datasets analysed during the current study are available in the Zenodo repository: https://doi.org/10.5281/zenodo.6458107.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Aktypi, A., Nurse, J.R.C., Goldsmith, M.: Unwinding Ariadne's identity thread: privacy risks with fitness trackers and online social networks. In: Proceedings of the 2017 on Multimedia Privacy and Security, pp. 1–11 (2017)

Al-Makhadmeh, Z., Tolba, A.: Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach. Measurement **147**, 106815 (2019)

Alhalabi, L., Singleton, M.J., Oseni, A.O., Shah, A.J., Zhang, Z.-M., Soliman, E.Z.: Relation of higher resting heart rate to risk of cardiovascular versus noncardiovascular death. Am. J. Cardiol. **119**(7), 1003–1007 (2017)

Alqhatani, A., Lipford, H.R.: Exploring the design space of sharing and privacy mechanisms in wearable fitness platforms. In: Workshop on Usable Security and Privacy (USEC), vol. 7 (2021)

Arca, S., Hewett, R.: Privacy protection in smart health. In: Proceedings of the 11th International Conference on Advances in Information Technology, pp. 1–8 (2020)

Bada, M., von Solms, B.: A cybersecurity guide for using fitness devices (2021). arXiv preprint arXiv:2105.02933

Balas, V.E., Solanki, V.K., Kumar, R., Ahad, M.A.R.: A Handbook of Internet of Things in Biomedical and Cyber Physical System. Springer (2020)

Becher, S., Gerl, A., Meier, B.: Don't forget the user: from user preferences to personal privacy policies. In: 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), pp. 774–778. IEEE (2020)

Blasco, J., Chen, T.M., Patil, H.K., Wolff, D.: Wearables security and privacy. In: Mission-Oriented Sensor Networks and Systems: Art and Science, pp. 351–380. Springer (2019)

Blow, F., Yen-Hung, H., Hoppa, M.: A study on vulnerabilities and threats to wearable devices. J. Colloq. Inf. Syst. Secur. Educ. **7**, 7 (2020)

Booth, F.W., Roberts, C.K., Laye, M.J.: Lack of exercise is a major cause of chronic diseases. Compr. Physiol. **2**(2), 1143 (2012)

Bourreau, M.: Google—Fitbit. https://voxeu.org/article/googlefitbit-will-monetise-health-data-and-harm-consumers (2020). Accessed 25 Dec 2021

Can a fitness tracker detect diabetes? https://precisiondrivenhealth.com/can-a-fitness-tracker-detect-diabetes/ (2017). Accessed 2 Aug 2022

CEOToday: is data the new gold? https://www.ceotodaymagazine.com/2018/04/is-data-the-new-gold/ (2020). Accessed 2 Aug 2022

Challa, N., Yu, S., Kunchakarra, S.: Wary about wearables: potential for the exploitation of wearable health technology through employee discrimination and sales to third parties. Intersect Stanford J. Sci. Technol. Soc. **10**(3) (2017)

Chang, L., Jiaqi, L., Wang, J., Chen, X., Fang, D., Tang, Z., Nurmi, P., Wang, Z.: Sleepguard: capturing rich sleep information using smartwatch sensing data. Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. **2**(3), 1–34 (2018)

Chen, L.F., Ismail, R.: Information technology program students' awareness and perceptions towards personal data protection and privacy. In: 2013 International Conference on Research and Innovation in Information Systems (ICRIIS), pp. 434–438. IEEE (2013)

Chen, Y., Shen, C.: Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access **5**, 3095–3110 (2017)

Cho, J.Y., Ko, D., Lee, B.G.: Strategic approach to privacy calculus of wearable device user regarding information disclosure and continuance intention. KSII Trans. Internet Inf. Syst. (TIIS) **12**(7), 3356–3374 (2018)

Cook, J.: Inferring religion. https://dzone.com/articles/inferring-personal-information-from-fitness-data (2021). Accessed 25 Dec 2021

Cooney, M.T., Vartiainen, E., Laakitainen, T., Juolevi, A., Dudina, A., Graham, I.M.: Elevated resting heart rate is an independent risk factor for cardiovascular disease in healthy men and women. Am. Heart J. **159**(4), 612–619 (2010)

Cremonini, M., Braghin, C., Ardagna, C.A.: Chapter 42—privacy on the internet. In: Vacca, J.R. (ed.), Computer and Information Security Handbook, 2 edn, pp. 739–753. Morgan Kaufmann, Boston (2013). ISBN: 978-0-12-394397-2. https://doi.org/10.1016/B978-0-12-394397-2.00042-8

Das, A.K., Pathak, P.H., Chuah, C.-N., Mohapatra, P.: Uncovering privacy leakage in BLE network traffic of wearable fitness trackers. In: Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications, pp. 99–104 (2016)

Dennedy, M.F., Fox, J., Finneran, T.R.: The Privacy Engineer's Manifesto: Getting from Policy to Code to QA to Value. Springer Nature (2014)

Fernström, M., Fernberg, U., Eliason, G., Hurtig-Wennlöf, A.: Aerobic fitness is associated with low cardiovascular disease risk: the impact of lifestyle on early risk factors for atherosclerosis in young healthy swedish individuals-the lifestyle, biomarker, and atherosclerosis study. Vasc. Health Risk Manag. **13**, 91 (2017)

Fietkiewicz, K., Ilhan, A.: Fitness tracking technologies: data privacy doesn't matter? the (un) concerns of users, former users, and non-users. In: Proceedings of the 53rd Hawaii International Conference on System Sciences (2020)
Forbes: data is the new gold. https://www.forbesafrica.com/technology/2019/07/18/data-is-the-new-gold/ (2019). Accessed 2 Aug 2022
Foukia, N., Billard, D., Solana, E.: Pisces: a framework for privacy by design in IoT. In: 2016 14th Annual Conference on Privacy, Security and Trust (PST), pp. 706–713. IEEE (2016)
Fourberg, N., Serpil, T., Wiewiorra, L., Godlovitch, Ilsa, De STreel, A., Jacquemin, H., Hill, J., Nunu, M., Jacques, F., Ledger, M., et al.: Online advertising: the impact of targeted advertising on advertisers, market access and consumer choice (2021)
Furberg, R., Brinton, J., Keating, M., Ortiz, A.: Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016 (2016). https://doi.org/10.5281/zenodo.53894
Gabriele, S., Chiasson, S.: Understanding fitness tracker users' security and privacy knowledge, attitudes and behaviours. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2020)
Gross, C., Wenner, W., Lackes, R.: Using wearable fitness trackers to detect covid-19? In: International Conference on Business Informatics Research, pp. 51–65. Springer (2021)
Hantke, F., Dewald, A.: How can data from fitness trackers be obtained and analyzed with a forensic approach? In: 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 500–508. IEEE (2020)
Henriksen, A., Mikalsen, M.H., Woldaregay, A.Z., Muzny, M., Hartvigsen, G., Hopstock, L.A., Grimsgaard, S., et al.: Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J. Med. Internet Res. **20**(3), e9157 (2018)
Hicks, J.L., Althoff, T., Sosic, R., Kuhar, P., Bostjancic, B., King, A.C., Leskovec, J., Delp, S.L.: Best practices for analyzing large-scale health data from wearables and smartphone apps. NPJ Digit. Med. **2**(1), 1–12 (2019)
Högström, G., Nordström, A., Nordström, P.: Aerobic fitness in late adolescence and the risk of early death: a prospective cohort study of 1.3 million Swedish men. Int. J. Epidemiol. **45**(4), 1159–1168 (2016)
Horvitz, E., Mulligan, D.: Data, privacy, and the greater good. Science **349**(6245), 253–255 (2015)
Hunter, S., Robson, S.C.: Adaptation of the maternal heart in pregnancy. Br. Heart J. **68**(6), 540 (1992)
Ilhan, A., Fietkiewicz, K.J.: Data privacy-related behavior and concerns of activity tracking technology users from Germany and the USA. Aslib J. Inf. Manag. **73**, 180–200 (2020)
IotaComm: how does iot affect big data? https://www.iotacommunications.com/blog/iot-big-data/ (2020). Accessed 27 June 2021
Jones, J.C., Seladi-Schulman, J.: Causes of slow heart rate. https://www.healthline.com/health/slow-heart-rate#causes (2021). Accessed 5 Nov 2021
Jung, G., Lee, H., Kim, A., Lee, U.: Too much information: assessing privacy risks of contact trace data disclosure on people with covid-19 in South Korea. Front. Public Health **8**, 305 (2020)
Kaiser, D.W., Harrington, R.A., Turakhia, M.P.: Wearable fitness trackers and heart disease. JAMA Cardiol. **1**(2), 239 (2016)
Kang, H., Jung, E.H.: The smart wearables-privacy paradox: a cluster analysis of smartwatch users. Behav. Inf. Technol. **40**(16), 1755–1768 (2021)
Kazlouski, A., Marchioro, T., Manifavas, H., Markatos, E.: Do you know who is talking to your wearable smartband? Integr. Citizen Centered Digit. Health Soc. Care Citizens Data Producers Serv. Co-Creators **275**, 142 (2020)
Kim, J.W., Moon, S.-M., Kang, S., Jang, B.: Effective privacy-preserving collection of health data from a user's wearable device. Appl. Sci. **10**(18), 6396 (2020)
Kounoudes, A.D.: Questionnaire on fitness trackers user privacy concerns. https://forms.gle/uzVzVhew2Jq3XeAS9 (2022a). Accessed 27 Mar 2022
Kounoudes, A.D.: PrivacyEnhaction Evaluation Questionnaire. https://forms.gle/KCJ2xx23quK4A8wk8 (2022b). Accessed 27 Mar 2022
Kounoudes, A.D., Kapitsaki, G.M.: A mapping of IoT user-centric privacy preserving approaches to the GDPR. Internet Things **11**, 100179 (2020)
Kounoudes, A.D., Kapitsaki, G.M., Katakis, I., Milis, M.: User-centred privacy inference detection for smart home devices. In: 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI), pp. 210–218. IEEE (2021)

Kröger, J.: Unexpected inferences from sensor data: a hidden privacy threat in the internet of things. In: IFIP International Internet of Things Conference, pp. 147–159. Springer (2018)
Kröger, J.L., Raschke, P., Bhuiyan, T.R.: Privacy implications of accelerometer data: a review of possible inferences. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy, pp. 81–87 (2019)
Krzanich, B.: Data is the new oil in the future of automated driving. https://newsroom.intel.com/editorials/krzanich-the-future-of-automated-driving/ (2016). Accessed 27 June 2021
Langley, M.R.: Hide your health: addressing the new privacy problem of consumer wearables. Geo. LJ **103**, 1641 (2014)
Lee, L., Lee, J., Egelman, S., Wagner, D.: Information disclosure concerns in the age of wearable computing. In: NDSS Workshop on Usable Security (USEC), vol. 1, pp. 1–10 (2016)
Lehto, M., Lehto, M.: Health information privacy of activity trackers. In: European Conference on Cyber Warfare and Security, pp. 243–251. Academic Conferences International Limited (2017)
Lovejoy, B.: Smartphone and smartwatch data led husband to confess to murdering his wife. https://9to5mac.com/2021/06/18/smartphone-and-smartwatch-data-murder/ (2021). Accessed 27 Mar 2022
Maganti, K., Rigolin, V.H., Sarano, M.E., Bonow, R.O.: Valvular heart disease: diagnosis and management. In: Mayo Clinic Proceedings, vol. 85, pp. 483–500. Elsevier (2010)
Michael Mangrum, J., DiMarco, J.P.: The evaluation and management of bradycardia. N. Engl. J. Med. **342**(10), 703–709 (2000)
Masuch, K., Greve, M., Trang, S.: Fitness first or safety first? Examining adverse consequences of privacy seals in the event of a data breach. In: Proceedings of the 54th Hawaii International Conference on System Sciences, p. 3871 (2021)
McGowan, E.: Here's what your Fitbit knows about you. https://blog.avast.com/what-fitbit-knows-about-you-avast (2021). Accessed 19 February 2022
Meteriz, Ü., Yıldıran, N.F., Mohaisen, A.: You can run, but you cannot hide: using elevation profiles to breach location privacy through trajectory prediction (2019). arXiv preprint arXiv:1910.09041
Mohzary, M., Tadisetty, S., Ghazinour, K.: A privacy protection layer for wearable devices. In: Foundations and Practice of Security: 12th International Symposium, FPS 2019, Toulouse, France, November 5–7, 2019, Revised Selected Papers, vol. 12056, p. 363. Springer Nature (2020)
Molich, R., Nielsen, J.: Improving a human–computer dialogue. Commun. ACM **33**(3), 338–348 (1990)
Nagai, M., Hoshide, S., Kario, K.: Sleep duration as a risk factor for cardiovascular disease-a review of the recent literature. Curr. Cardiol. Rev. **6**(1), 54–61 (2010)
Obar, J.A., Oeldorf-Hirsch, A.: The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. Inf. Commun. Soc. **23**(1), 128–147 (2020)
Pan, S.B.: Get to know me: protecting privacy and autonomy under big data's penetrating gaze. Harv. JL Tech. **30**, 239 (2016)
Parate, A.: Designing efficient and accurate behavior-aware mobile systems (2014)
Peek, S.T.M., Wouters, E.J.M., Van Hoof, J., Luijkx, K.G., Boeije, H.R., Vrijhoef, H.J.M.: Factors influencing acceptance of technology for aging in place: a systematic review. Int. J. Med. Inform. **83**(4), 235–248 (2014)
Peppet, S.R.: Regulating the internet of things: first steps toward managing discrimination, privacy, security and consent. Tex. L. Rev. **93**, 85 (2014)
Perez, A.J., Zeadally, S., Cochran, J.: A review and an empirical analysis of privacy policy and notices for consumer internet of things. Secur. Privacy **1**(3), e15 (2018)
Prince, A.: Location as health. Houston Journal of Health Law and Policy, Forthcoming, U Iowa Legal Studies Research Paper (2021-06) (2021)
Psychoula, I., Chen, L., Amft, O.: Privacy risk awareness in wearables and the internet of things. IEEE Pervasive Comput. **19**(3), 60–66 (2020)
Rahmany, M., Zin, A.M., Sundararajan, E.A.: Comparing tools provided by python and r for exploratory data analysis. Int. J. Inf. Syst. Comput. Sci. (IJISCS) **4**(3), 131–142 (2020)
Reichherzer, T., Timm, M., Earley, N., Reyes, N., Kumar, V.: Using machine learning techniques to track individuals & their fitness activities. In: CATA 2017, pp. 119–124. ISCA (2017)
Reinhardt, D., Borchard, J., Hurtienne, J.: Visual interactive privacy policy: the better choice? In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–12 (2021)
Saint-Maurice, P.F., Troiano, R.P., Bassett, D.R., Graubard, B.I., Carlson, S.A., Shiroma, E.J., Fulton, J.E., Matthews, C.E.: Association of daily step count and step intensity with mortality among us adults. JAMA **323**(12), 1151–1160 (2020)

Sarstedt, M., Mooi, E.: Descriptive statistics. In: A Concise Guide to Market Research, pp. 91–150. Springer (2019)
Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Arora, T., Taheri, S.: Sleep quality prediction from wearable data using deep learning. JMIR Mhealth Uhealth **4**(4), e125 (2016)
Sigmund, T.: Attention paid to privacy policy statements. Information **12**(4), 144 (2021)
Skiljic, A.: Health inferences. https://iapp.org/news/a/the-status-quo-of-health-data-inferences/ (2021). Accessed 5 Nov 2021
Tang, Q.: Automated Detection of Puffing and Smoking with Wrist Accelerometers. Northeastern University (2014)
Tedesco, S., Sica, M., Ancillao, A., Timmons, S., Barton, J., O'Flynn, B.: Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. PLoS ONE **14**(5), e0216891 (2019)
Thakkar, P.K., He, S., Xu, S., Huang, D.Y., Yao, Y.: "It would probably turn into a social faux-pas": users' and bystanders' preferences of privacy awareness mechanisms in smart homes. In: CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2022)
Thomaz, E., Essa, I., Abowd, G.D.: A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 1029–1040 (2015)
Torre, I., Koceva, F., Sanchez, O.R., Adorni, G.: Fitness trackers and wearable devices: how to prevent inference risks? In: Proceedings of the 11th EAI International Conference on Body Area Networks, pp. 125–131 (2016)
Tudor-Locke, C., Bassett, D.R.: How many steps/day are enough? Sports Med. **34**(1), 1–8 (2004)
Vailshery, L.S.: IoT connected devices worldwide 2030. https://www.statista.com/statistics/802690/worldwide-connected-devices-by-access-technology/ (2021). Accessed 27 June 2021
Valdez, P.: Focus: attention science: circadian rhythms in attention. Yale J. Biol. Med. **92**(1), 81 (2019)
Velykoivanenko, L., Niksirat, K.S., Zufferey, N., Humbert, M., Huguenin, K., Cherubini, M.: Are those steps worth your privacy? Fitness-tracker users' perceptions of privacy and utility. Proc. ACM Interact. Mobile Wearable Ubiquitous Technol. **5**(4), 1–41 (2021)
Vemou, K., Karyda, M., Kokolakis, S.: Directions for raising privacy awareness in SNS platforms. In: Proceedings of the 18th Panhellenic Conference on Informatics, pp. 1–6 (2014)
Vitak, J., Liao, Y., Kumar, P., Zimmer, M., Kritikos, K.: Privacy attitudes and data valuation among fitness tracker users. In: International Conference on Information, pp. 229–239. Springer (2018)
Vuori, I.: Physical inactivity is a cause and physical activity is a remedy for major public health problems. Kinesiology **36**(2), 123–153 (2004)
Wachter, S., Mittelstadt, B.: A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Column Bus Law Rev. **2019**, 494 (2019)
Webster, D.E., Tummalacherla, M., Higgins, M., Wing, D., Ashley, E., Kelly, V.E., McConnell, M.V., Muse, E.D., Olgin, J.E., Mangravite, L.M., et al.: Smartphone-based vo2max measurement with heart snapshot in clinical and real-world settings with a diverse population: Validation study. JMIR Mhealth Uhealth **9**(6), e26006 (2021)
WEF: Data is the new gold. This is how it can benefit everyone—while harming no one. https://bit.ly/3eazKmm (2020). Accessed 2 Aug 2022
Whittaker, Z.: How Strava's "anonymized" fitness tracking data spilled government secrets. https://www.zdnet.com/article/strava-anonymized-fitness-tracking-data-government-opsec/ (2018). Accessed 17 Feb 2022
Wu, Q., Sum, K., Nathan-Roberts, D.: How fitness trackers facilitate health behavior change. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 60, pp. 1068–1072. SAGE Publications Sage CA, Los Angeles (2016)
Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., Chen, Z.: How much can behavioral targeting help online advertising? In: Proceedings of the 18th International Conference on World Wide Web, pp. 261–270 (2009)
Yan, T., Lu, Y., Zhang, N.: Privacy disclosure from wearable devices. In: Proceedings of the 2015 Workshop on Privacy-Aware Mobile Computing, pp. 13–18 (2015)
Yao, Y., Song, L., Ye, J.: Motion-to-BMI: using motion sensors to predict the body mass index of smartphone users. Sensors **20**(4), 1134 (2020)

Zimmer, M., Kumar, P., Vitak, J., Liao, Y., Kritikos, K.C.: There's nothing really they can do with this information: unpacking how users manage privacy boundaries for personal fitness information. Inf. Commun. Soc. **23**(7), 1020–1037 (2020)

Alexia Dini Kounoudes is currently pursuing a Ph.D. degree and is a member of the Software Engineering and Internet Technologies (SEIT) Lab of the Department of Computer Science at the University of Cyprus. She received her MSc in Software Engineering from the University of Bradford, UK (1999), and her BSc in Computer Science from the University of Kent at Canterbury, UK (1998). She has 24 years of industrial experience in various positions in the software development sector both in Cyprus and the UK. Her research interests include Privacy-Enhancing Technologies, User Privacy, User Privacy in IoT, User Privacy Protection under the GDPR.

Georgia M. Kapitsaki is an Associate Professor in the Department of Computer Science at the University of Cyprus. She received her Ph.D. in Electrical and Computer Engineering from the National Technical University of Athens, Greece, in 2009. She has been involved in the organisation of international conferences (e.g. ICSME 2022, SAC 2019, ICSR 2016, etc.). She has been serving as a member of the program committee of international conferences (e.g., WISE 2022, ICWS 2022, MSR 2022, ENASE 2021, WEBIST 2021). She has published over 60 papers in international journals and conferences, and is contributing to open source projects. She has received the best paper award in ICSR 2015 and in the doctoral symposium of MODELS 2008. She has served as an evaluator for EU proposals. She has been involved and received funding from European research projects (e.g. SocioCoast, CYberSafety, TAMIT). Her research interests include Software Engineering, Open Source Software, Human Aspects in Software Engineering, Context-aware applications and Privacy Enhancing Technologies.

Ioannis Katakis is a Faculty Member at the University of Nicosia. He studied Computer Science and holds a PhD in Machine Learning for Text Classification. After his post-graduate studies, he served various universities as a lecturer and a senior researcher (Aristotle University of Thessaloniki, University of Cyprus, Cyprus University of Technology, Open University of Cyprus, Hellenic Open University, Athens University of Economics and Business, National and Kapodistrian University of Athens). His research interests include Mining Social, Web and Urban Data, Sentiment Analysis and Opinion Mining, Data Streams, Multi-label Learning. He has published papers in International Conferences and Scientific Journals related to his areas of expertise (e.g. CIKM, ECML/PKDD, IEEE TKDE, ECAI), organized three workshops (at ICML, ECML/PKDD, EDBT/ICDT), edited four special issues (DAMI, InfSys) and is an Editor at the journal Information Systems. His research has been cited more than 4200 times in the literature. He regularly serves the programme committee of conferences (ECML/PKDD, WSDM, DEBS, IJCAI) and evaluates articles in numerous journals (TPAMI, DMKD, TKDE, TKDD, JMLR, TWEB, ML).