

**A Sociological Approach to User Privacy in the Internet of  
Things and Smart City Environments**

by

**Srihaasa Pidikiti**

B.Tech, Computer Science and Engineering, 2018

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Masters  
Department of Computer Science  
2021

Committee Members:  
Shivakant Mishra, Chair  
Celeste Campos-Castillo  
Tamara Lehman

Pidikiti, Srihaasa (M.S., Computer Science)

A Sociological Approach to User Privacy in the Internet of Things and Smart City Environments

Thesis directed by Prof. Shivakant Mishra

With the emergence of the Internet of Things in a broad range of sectors, there is a looming threat to user privacy due to the continuous activity monitoring or tracking and the limited control over the data collected. Users are expected to assume collective responsibility of safeguarding privacy with the service providers by specifying their privacy preferences. Nonetheless, users massively fail at this task due to the precariousness associated with IOT applications functioning and their limited cognitive abilities. The main goal of this work is to develop a privacy preference prediction mechanism to assist the users in their decision-making process considering demographic, social media activity and contextual factors. We conduct a survey with 240 participants across three continents on Amazon Mechanical Turk to identify the significant factors that influence an individual's privacy choices. We further use these recognized factors alongside the associated familiarity of the user with the given IOT scenario to design a privacy preference rating classification model and achieve an accuracy of 72% in our endeavor. With this privacy classifier as the backbone, we propose a personalized novel web application architecture to provide users with information available concerning their privacy. Our work is the first of its kind that leverages a combination of demographic, social media activity and contextual factors in determining privacy preferences.

## Contents

<b>Chapter</b>	
<b>1</b>	<b>Introduction . . . . . 1</b>
<b>2</b>	<b>Literature Review . . . . . 4</b>
2.1	Privacy Significance in IOT . . . . . 4
2.2	Factors influencing Privacy Preferences . . . . . 5
2.3	Privacy Preferences Prediction . . . . . 7
2.4	Privacy Awareness . . . . . 8
<b>3</b>	<b>Methodology . . . . . 10</b>
3.1	Survey Design . . . . . 10
3.2	Privacy Preference Classification . . . . . 14
3.3	Privacy Awareness Assistant . . . . . 15
<b>4</b>	<b>Results . . . . . 17</b>
4.1	Demographic Factors . . . . . 18
4.2	Social Media Activity Factors . . . . . 21
4.3	Contextual Factors . . . . . 23
4.4	Additional Miscellaneous Factors . . . . . 25
4.5	Qualitative Analysis . . . . . 26
4.6	Privacy Prediction . . . . . 30

4.6.1	XGBoost Classifier . . . . .	31
4.6.2	SVM Classifier . . . . .	33
4.7	Privacy Awareness Assistant . . . . .	33
<b>5</b>	<b>Discussion</b>	<b>39</b>
5.1	Limitations . . . . .	40
<b>6</b>	<b>Conclusion</b>	<b>42</b>
	<b>Bibliography</b>	<b>43</b>
	<b>Appendix</b>	

## Tables

### Table

3.1	Sample IOT scenarios presented to survey participants . . . . .	14
3.2	List of examined IOT applications . . . . .	16
4.1	Privacy Concerns of participants in Static Entertainment IOT scenarios . . . . .	28
4.2	Privacy Concerns of participants in Dynamic Entertainment IOT scenarios . . . . .	32
4.3	Privacy Concerns of participants in Static Financial IOT scenarios . . . . .	34
4.4	Privacy Concerns of participants in Dynamic Financial IOT scenarios . . . . .	38

## Figures

### Figure

4.1	Average usage ratings of the participants grouped by Educational Qualifications . . .	18
4.2	Average usage ratings of the participants grouped by Continent . . . . .	19
4.3	Average usage ratings of the participants grouped by Ethnicity . . . . .	20
4.4	Average usage ratings of the participants grouped by Marital Status . . . . .	21
4.5	Average usage ratings of the participants grouped by Religion . . . . .	22
4.6	Average usage ratings of the participants grouped by Age and Continent . . . . .	23
4.7	Average usage ratings of the participants clustered by Social media activity score . .	24
4.8	Average usage ratings of the participants based on the scenario's context . . . . .	25
4.9	Average usage ratings of the participants based on their familiarity with IOT devices	26
4.10	Average usage ratings of the participants based on their familiarity with the given IOT scenario . . . . .	27
4.11	Questionnaire about demographic information in Privacy Awareness Assistant . . . .	35
4.12	Questionnaire about social media activity in Privacy Awareness Assistant . . . . .	36
4.13	Questionnaire about IOT application in Privacy Awareness Assistant . . . . .	37
4.14	Questionnaire about privacy concern in Privacy Awareness Assistant . . . . .	38

## **Chapter 1**

### **Introduction**

In recent years, IOT based applications and devices have successfully penetrated day-to-day activities making our lives easier. According to the IDC forecast report, [14], there will be 55.7 billion IOT connect devices worldwide by the year 2025. A wide range of IOT devices is available in the market to cater to several everyday scenarios encountered by users. Most of these devices gather personal and identifiable information to facilitate end-users with high-quality personalized services. Users can be unaware of this plethora of information that is being collected in the background. Besides, there is an imminent risk with this information being sold to third parties by the IOT service providers. Moreover, nowadays various privacy endangering events like the misuse of shared information, malicious malware attacks, ineffective user data encryption [56], [62] are being reported by the media. These news reports can further aggravate the privacy concerns of IOT users and ultimately result in them being less likely to use such IOT applications. As a consequence, safeguarding the user's privacy in the IOT environment is of utmost importance to ensure a hassle-free adoption of IOT.

Preserving the privacy of its users presents a mammoth challenge to the service providers. A combination of privacy-enhancing technologies and legal means are needed by service providers to protect the privacy of their users [50]. Nevertheless, privacy is more of an individual-centric concept. Thus, service providers tend to share the burden of managing privacy by taking into account the privacy preferences of users like consent for data sharing with third parties. But, users, in general, tend to fail at making precise privacy choices owing to uncertainty, malleability,

misdeeming integrational advantages and cognitive impairment [15], [58].

Many solutions have been proposed by the researchers over the years to predict privacy preferences of IOT users based on the impact of contextual factors and the user interaction history [37], [17]. Although these works aimed at assisting the users with the complex task of deciding privacy settings, there is no prior work that builds prediction models considering the combined influence of demographic, social media activity and contextual factors - each of which previous research links to privacy preferences - in IOT environments. Our novel contribution is to systematically explore the impact of demographic factors like age, gender, religion, ethnicity, continent, education and marital status alongside social media activity and contextual factors on the privacy choices of an individual. Mainly, our objective is to design a privacy prediction mechanism based on the answers to the following research questions:

- **RQ1:** Do demographic factors have a significant impact on privacy preferences in IOT applications?
- **RQ2:** Does the social media activity of IOT users have an important role in privacy preferences?
- **RQ3:** How do the contextual parameters, level of user control over information collection and utilization purpose influence privacy choices?
- **RQ4:** What is the role of familiarity with similar IOT applications, when deciding on privacy settings?

In order to answer these questions, we leverage the Amazon Mechanical Turk platform to administer a survey designed to target a diverse sample. We conduct surveys across three continents- North America, Europe and Asia and collect responses of 240 participants in total. We present each participant with 12 different IOT applications that constitute a balanced set in terms of the level of user control and purpose of utilization. For each of these applications, we ask the participants to rate the likeliness to use it and the degree of familiarity with similar applications.



Subsequently, we study the impact of demographic, social media and contextual factors on these choices of participants. Our study makes the following findings:

- We show that Education, Ethnicity, Continent, Religion, Marital status and the calculated Social Media activity score have a significant impact on privacy choices regarding IOT applications.
- We observe that the level of control given by the IOT application to the user (static vs dynamic) and utilization purpose (finance vs entertainment) change the user’s privacy perceptions.
- We identify that familiarity with similar IOT application scenarios and the ultimate privacy decisions are highly correlated with each other.
- Finally, we conduct a qualitative survey to identify the primary privacy concerns in using IOT applications generalized in terms of the level of user control and the purpose of utilization. We further use the findings from the analysis of these responses to corroborate the importance of certain factors in making privacy-related decisions.

Lastly, we take into consideration the above-mentioned findings regarding demographic, social media activity and contextual factors to build a machine learning-based privacy classification model. We achieve an accuracy of 72% using this model in predicting the privacy preference range of individuals for a given context. Building upon this privacy classifier, we suggest an architecture for a web-based application that further assists IOT users in determining their likeliness to use the IOT application by providing obtainable information regarding an individual’s specific privacy concerns.

## **Chapter 2**

### **Literature Review**

In this chapter, we discuss the prior work done in identifying the importance of privacy in IOT environments and the existing mechanisms employed in protecting the privacy of the users. To aid our goal to devise a new privacy preference prediction mechanism, we look into previous works of researchers that identify significant demographic and contextual factors that impact privacy preferences. The majority of these studies complement our work and further confirm our findings.

#### **2.1 Privacy Significance in IOT**

With the unprecedented use of IOT devices worldwide and the plethora of information collected by these devices regularly, safeguarding the privacy of the users has become a major concern. Mishandled privacy protection poses significant threats to the sensitive identifiable information of the users like User identification, User Tracking, Profiling, Utility monitoring and controlling [54]. In addition to the possible privacy invasion of information from external entities according to an article [20] there have been reports of abuse or domestic violence using the smart home devices by monitoring partner's activities. Moreover, a prior study [16] reveals the negative impact of privacy concerns on the trustworthiness of an IOT application.

IOT service providers need to consider all these privacy concerns and address them accordingly to gain the user's trust. However, Lee et.al. suggests that privacy preservation is a counterproductive task to the IOT service providers due to the potential use of the generated data to

increase the quality of the IOT service and decrease the costs incurred by the providers [39]. Comprehending all these findings regarding privacy, a trade-off solution between IOT service providers and users has to be formulated. Our work proposes an efficient sociological and contextual approach to tackle privacy concerns.

## 2.2 Factors influencing Privacy Preferences

Several studies have looked into the impact of demographic differences on users' privacy preferences in general on the internet. Citizens in countries with a higher ranking in uncertainty avoidance are more likely to emphasize privacy risks in social network sites like Facebook and Twitter [61]. One previous study [43] identifies that mainly people in individualistic countries are more apprehensive about information privacy, unlike collectivistic countries where there is a common acceptance of privacy invasion by the organizations upon which the population is emotionally dependent. Older adults and women within individualistic countries have stronger internet privacy concerns in comparison to their peers [24]. Further, a survey conducted in the U.S found that people with higher educational qualifications like doctoral degrees have more privacy concerns, unlike the others [46]. Our work primarily focuses on the impact of such factors in the IOT domain specifically. We also discuss the significant effect of additional demographic factors, marital status and religious beliefs on the privacy concerns in the IOT environment.

Contextual parameters have a notable contribution in the process of setting a user's privacy preferences in IOT applications. Based on the survey results conducted with five contextual parameters- entity collecting data, type, location, reason and persistence of data collection, it is discerned that users had major privacy concerns if the data type is either an audio or video and has a possible government access [17]. Using the same parameter definitions, another study reveals that users tend to have issues with data gathering at private places rather than the public places [37]. This work also identifies greater privacy concerns with uninterrupted data collection. With respect to the type of data involved, users are less comfortable when it comes to sharing banking details compared to exercise data or shopping details [51]. Users are more cautious about their

privacy when the retention period of the data is beyond a week [44]. Our work further extends on these findings considering a completely new contextual parameter that emphasizes the level of user control on data collection- whether it is explicitly given and updated by the user or collected and inferred in the background. Furthermore, we give priority to two IOT application utilization purposes in particular which are financial and entertainment in conjunction with the contextual parameter, level of user control to analyze the differences in privacy preferences.

Location is one of the contributing factors in privacy preference settings. Extensive research work has been done to understand the privacy concerns of smart home locations in particular. Zheng et.al. explains that convenience attained in performing regular household activities on using a smart home device diminishes the privacy concerns involved [63]. Smart home users are specifically concerned about the recordings of self-appearance, intimacy, cooking, eating, media use and oral expressions[25]. Prior internet expertise also plays a vital role in deciding privacy preferences. Bellman et.al. concluded that the users with relatively more internet experience exhibit fewer privacy concerns compared to their counterparts with low internet experience [18]. In contrast, another research work [57] explains that greater internet expertise causes higher internet privacy concerns due to their knowledge of the internet. Our work is a novel attempt that studies the impact of social media expertise determiners as the number of followers, usage duration, post or tweet privacy settings, etc. on the IOT privacy preferences of an individual.

Some prior studies conducted qualitative surveys to corroborate their quantitative survey findings and to make better inferences of an individual's privacy concerns based on their responses. One such study [63] found that people had contrasting opinions about smart home devices sharing data with advertisers. Some of them felt that this can help them in getting good recommendations whereas others felt the need for transparency with the shared data. Another study [51] questioned the survey participants about their privacy concerns in certain IOT scenarios. Our contribution is the shift of qualitative analysis from these scenario-based questions to broader questions about four predefined contextual categories based on two parameters, level of user control and utilization purpose - static finance, static entertainment, dynamic finance and dynamic entertainment.

## 2.3 Privacy Preferences Prediction

Setting privacy preferences can be a daunting task for users. Many researchers have proposed solutions to alleviate this issue by making predictions of user's privacy preferences. Sadeh et.al. proposed one such solution for setting privacy preferences of a mobile application that selectively shares a location with other requesters based on time and requester type [53]. This work performs a supervised classification using Random Forest to refine user's privacy settings based on their prior privacy choices. On a whole, this study shows an increase of accuracy from 79% with user-defined preferences to 91% with classifier-defined preferences. The authors explained the rationale behind this observation that users are not highly successful in determining their privacy preferences beforehand which are consistent in long run with their actual location sharing choices. In a similar context of location sharing application, another study [52] used decision tree and clustering techniques to define canonical privacy policies that helped reduce the burden on users.

Fang et.al. built a machine learning model that achieves more than 90% accuracy in predicting privacy preferences on Facebook with the help of user-defined labels on sharing data with selective friends [29]. Similarly, to predict personalized privacy preferences on social media, another study [28] used J48 decision models that take into account psychological factors like trustworthiness, sensitivity and appropriateness besides the conventional contextual factors. All these above discussed works mainly implemented privacy prediction models for social media or mobile networks. Our work on the other hand makes privacy predictions exclusively for IOT based applications.

Lee et.al. defined four clusters mainly based on the contextual parameters- who collects the data and the type of data involved, to predict privacy preferences in IOT scenarios with about 77% accuracy.[37]. Extending on this work, Bahirat et.al. defined clusters at the user level rather than the scenario level that are further used to build default smart profiles. This work uses a combination of clustering, tree-learning algorithms and predicts privacy preferences with an accuracy of 82% [17]. Another study established that general privacy segmentation of users along with contextual parameters helps to predict people's privacy behavior more accurately [38]. Besides taking into

account contextual parameters, our work includes a wide range of demographic parameters like the continent, religion, marital status, age and social media usage for designing privacy preference prediction models. Researchers found that including demographic factors like country, language and individualism index improves the accuracy of privacy prediction models [40]. Our contribution enhances the findings of this prior work by looking at a more comprehensive list of demographic factors like age, gender, religion, education etc.

## 2.4 Privacy Awareness

Data privacy laws like Federal Trade Commission Act (FTC), California Consumer Privacy Act (CCPA), General Data Protection Regulation (GDPR) require IOT service providers to provide information to users about the practices they follow to safeguard user’s privacy. However, such information providing lengthy privacy policies are generally not efficient in informing the users due to their complexity. Users tend to skip reading the entire policy when presented with an option to sign the agreement without reading [59]. A prior effort demonstrated that simple and concise privacy policies contributed to a better understanding of the data practices [55]. Gluck et.al reconfirms in their work that condensing lengthy privacy policies into succinct privacy notices increases awareness about privacy [30]. But further reducing these notices to include certain data practices led to a decrease in awareness of the excluded practices. Taking advantage of these findings, our proposed personalized web application provides users with brief information extracted from a privacy policy that is relevant to their mentioned privacy concerns.

Various privacy awareness systems are designed that inform users about their personal information collection and further control data collection based on user’s privacy preferences. One such system is pawS [35] which uses machine-readable XML privacy policy like P3P [26] given by service providers and privacy proxies defined by users that are configured using a preferred language such as APPEL. But in recent years since these P3P privacy policies are no longer widely adopted, service providers might not prefer to change their infrastructure to suit this system’s requirements. To remove any such dependency on service providers, Kolter et.al proposes a collaborative system

[33] with a community that allows experienced users to write and share an understandable explanation of a provider’s privacy policy. Inexperienced users can benefit from this shared information about the application and can even import expert’s privacy preferences. These systems deal with website-based privacy mainly whereas our system solely focuses on IOT based privacy.

Das et.al designed an IOT based personalized privacy assistant that provides information about data practices of all the nearby IOT devices [27]. This assistant which is a smartphone application fetches privacy information about the IOT application provided in IOT Resource Registry (IRR) by the service providers. With the help of this assistant users can even configure privacy settings for the various IOT applications when such settings are available. Based on this proposed system architecture, an IOT privacy research framework named TIPPERS [42] is developed and deployed in a smart building scenario [47]. This system includes privacy-aware smart buildings which publish building privacy policies through IRRs, fetch user’s privacy preferences from IOT assistant which is a smartphone application and enforce them during data collection/ sharing with IOT service providers. Lee et.al designed a web-based privacy awareness system called IOT Service Store [36] which provides users information about the possible inferences that can be made from their shared personal information. This system also includes a rating system where users can collaboratively evaluate the tradeoff between utility and privacy of an IOT application.

Our proposed system taking into consideration that users might not always be able to articulate explicitly their privacy preferences beforehand makes personalized predictions about their privacy preferences in terms of their likeliness to use the IOT application. This prediction in addition to the information about general privacy concerns can help even inexperienced users to determine whether or not to use the IOT application. Moreover, our system is not dependent on direct collaboration with service providers, instead, users can upload the terms of agreement given by service providers at the time of installation to learn more about the information regarding their specific privacy concerns if any.

## **Chapter 3**

### **Methodology**

We design the survey and choose a sufficient sample size specifically aimed to address the research questions RQ1, RQ2, RQ3 and RQ4. Further, we analyze the responses and develop a privacy prediction mechanism based on our findings. The experimental protocol for this survey is reviewed and conducted under IRB Protocol 20-0530.

#### **3.1 Survey Design**

We utilize the Amazon Mechanical Turk platform to conduct our designed survey across three continents Asia, Europe and North America. We collect demographic information of participants like Age, Gender, Continent they currently reside in, Ethnicity/ Race, Religious views they follow, Educational background and Marital status. We pay \$2 to each participant for successful completion of the survey. To ensure the overall quality of the survey, we include attention-check questions. In total, we have 240 participants who attempted our survey and passed the quality check, 80 from each of the three continents- Asia, Europe and North America. We have an equal distribution of participants 50% each from the age groups, 18-30 and 31-50. Similarly, we have 50% (120 participants) who identified themselves as Female and another 50% (120 participants) as Male. The majority of our participants describe their ethnicity as either Asian/ Pacific Islander (38%, 90 participants) or White (56%, 134 participants). With respect to the religious views, 35% (85 participants) follow Hinduism, 26% (62 participants) follow Christianity whereas 27% (65 participants) are non-religious.



We categorize the educational backgrounds of participants into three tiers- Tier 1, Tier 2 and Tier 3. Tier 1 represents participants who have one of the following educational qualifications- Trade/ technical/ vocational training, Associate/ 2-year degree, High School degree or equivalent, Less than a high school diploma. Tier 2 represents participants who have finished either a Bachelor's or 4-year college degree. Lastly, Tier 3 represents participants with either a Master's degree or a Doctorate. In total, 25% (61 participants) belong to Tier 1, 48% (115 participants) belong to Tier 2 and 27% (64 participants) belong to Tier 3 respectively. Most of our participants are either married (50%, 121 participants) or single (49%, 117 participants). Information gathered from the participants about these demographic identifiers helps in answering the first research question:

**RQ1:** *Do demographic factors have a significant impact on privacy preferences in IOT applications?*

Besides collecting demographic information of the participants, we asked the participants to answer a few questions regarding their social media activity like the number of followers or friends, time spent on social media in a week, default privacy settings for posts on social media. Based on the responses, we find that 15 of the total participants have less than or equal to 10 followers, 25 participants have 11 to 50 followers, 23 participants have 51 to 100 followers, 40 participants have 101 to 200 followers, 39 participants have 201 to 300 followers, 23 participants have 301 to 400 followers and our distribution has the highest number of participants, 73 in number, with more than 400 followers. Regarding the time spent by our participants on social media in a week, a very less number of eight participants in our study, spend less than 10 minutes. 35 participants spend 10 to 30 minutes, 48 participants spend 31 to 60 minutes, 59 participants spend one to two hours, 37 participants spend two to three hours and 51 participants spend more than three hours.

Most of our participants about 124 in total have their default privacy settings as Friends only, 56 of them have public privacy settings, 40 have privacy set as Friends of Friends whereas 18 have it as Selected Friends only. We also take opinions of participants in general about social media by collecting responses about participants' agreement with the following two statements:

**Statement 1:** *Social media is a part of my everyday activity.*

**Statement 2:** *I would be sorry with a social media shut down.*

We find that a greater share of our participants either strongly agreed (85 participants) or agreed (125 participants) with the first statement that social media browsing occupies a portion of their everyday time. We observe similar behavior with the agreement ratings for the second statement also. Mainly participants, either strongly agreed (85 participants) or agreed (125 participants) that cease of social media operation will affect them. We do not have any representation of participants who strongly disagreed with any of the two statements. However, concerning the first statement, we do have a few participants who are neutral in terms of agreement (13 participants) and disagreed (9 participants). We notice a slight increase in the number of participants who are neutral in terms of agreement (33 participants) and disagreed (25 participants) for the second statement.

Lastly, in our attempt to understand the familiarity of our participants with the Internet of Things devices in general, we require participants to rate themselves on a scale of familiarity from one to five, one being not familiar and five being very familiar with IOT. Interestingly most of our survey participants are familiar enough and only a small number of 16 participants are completely unfamiliar with IOT devices. Collectively, all these social media activity identifiers of the participants help in answering the second research question:

**RQ2:** *Do social media activity factors have an important role in privacy preferences for IOT applications?*

We present the participants with 12 hypothetical IOT scenarios that systematically vary two contextual parameters. To generate these scenarios, we define the first contextual parameter as the level of user control on the data collection. We classify user control of information collection into two discrete types- static and dynamic. Static control refers to the collection of information that is explicitly given or updated by the user like age, fingerprint, gender, passwords, song playlist etc. Dynamic control refers to the information collection or inference in the background about the user by the application like location, activity tracking etc. This contextual parameter aims at resolving the third research question:

**RQ3:** *How does the contextual parameter, level of user control over information collection, influence privacy choices?*

The second contextual parameter identifies the primary purpose of utilization of the IOT application discussed in the scenario. We specifically choose scenarios about IOT applications that have either a financial or entertainment utilization purpose to them. Our thought process behind this choice is to ensure that our study includes the extreme ends of information sensitivity in the IOT spectrum.

We conduct our survey using a set of scenarios balanced in terms of the two chosen contextual parameters- User control and Purpose. Altogether, as shown in Table 3.1, we have four contextual-based categories, which are Dynamic- Financial, Dynamic- Entertainment, Static- Financial, Static- Entertainment. For our survey purpose, we consider two scenarios of each of these categories. Additionally, the survey also contains two miscellaneous purpose scenarios of both levels of user control, static and dynamic. For each of these 12 scenarios shown to the participants, we ask them to provide their likelihood and familiarity ratings as responses to the following four questions:

**Question 1:** *How familiar are you with such applications?*

**Question 2:** *How likely are you to use this application?*

**Question 3:** *How likely are you to use this application if your family or your friends are using it?*

**Question 4:** *How likely are you to use the application if a million people use it?*

For Question 1, we give a familiarity rating scale increasing from one to five with one being not at all familiar and five being extremely familiar. All the remaining three questions, Question 2, Question 3 and Question 4 have the same rating scale of likeliness to use similar IOT applications as discussed in the scenario. This likeliness rating scale ranges from one to five, with one being very unlikely and five being very likely. Responses to Question 1 assist in our attempt to answer the fourth research question:

**RQ4:** *What is the role of familiarity with similar IOT applications, when deciding on privacy settings?*

Responses to questions two, three and four about the likeliness to use IOT applications help in forming an understanding of privacy preferences of participants and the impact of demographic, social media activity and contextual factors in this regard. Lastly, at the end of the survey, we ask the participants to mention their specific privacy concerns in using IOT applications that have either static or dynamic control over information collection and are used either for financial or entertainment purposes.

Table 3.1: Sample IOT scenarios presented to survey participants

Control level	Purpose	Sample Scenarios
Dynamic	Financial	<i>An app that keeps track of your spending habits and recommends discounts based on your purchase history. This app stores information about your credit card details and collects transactions performed using it.</i>
Dynamic	Entertainment	<i>A personalised music app that allows you to save songs that are to be played when doing particular activities during the day like jogging, meditation, dancing etc. The app detects the activity being performed and plays the saved music associated with the activity.</i>
Static	Financial	<i>An app where you can save details about all of your mobile bank accounts in one place, including your login and password information, so that you can access all of them by using just this app.</i>
Static	Entertainment	<i>A shopping app wants you to provide your personal information like gender, date of birth, height and weight to provide you personal recommendations about the clothes you may purchase that are well suited to you.</i>

### 3.2 Privacy Preference Classification

Initially, we investigate the significance of the impact of the various demographic factors like Age, Gender, Religion, Education etc. chosen in our study on the privacy preferences of participants for the given IOT scenarios. Further, we cluster participants based on responses about their social media activity. We employ the k-means algorithm [34] for this purpose and the value of k is chosen using the elbow method. We use this cluster identifier as the social media activity

representation of the participant. Finally, using significant demographic factors, cluster identifier of social media activity, familiarity with IOT in general and contextual parameters as the features, we design privacy preference classification models. These models are built using supervised machine learning techniques. Specifically, we develop Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) classifiers and have chosen the best performing model based on performance metrics.

### **3.3 Privacy Awareness Assistant**

We examine the available privacy policies or terms of service agreement for 13 IOT applications in total to identify the common data practices and privacy concerns being addressed. We then design our web-based privacy assistant to annotate the provided policies and fetch, store information about such common concerns to answer future queries of users. We choose this mixed representation of 13 IOT applications as shown in 3.2 based on their possible utilization purpose. Each of these applications falls into one or more of the three utilization categories- entertainment, financial, smart home. We specifically included entertainment and financial categories as these have been already chosen in the earlier part of this work owing to their extreme information sensitivities. In addition to these two categories, we also inspect smart home privacy policies due to the high relevance of smart home applications in the market.

Table 3.2: List of examined IOT applications

<b>Name</b>	<b>Purpose</b>
Alexa [1]	Financial, Entertainment, Smart home
Garmin Pay [4]	Financial
Apple TV [2]	Entertainment
Samsung Family Hub Refrigerator [10]	Financial, Entertainment
SimpliSafe [11]	Smart home
Xfinity Voice Remote [13]	Entertainment
Honeywell Total Connect Comfort [6]	Smart home
Wemo WiFi Smart Plug [12]	Smart Home
Fitbit [3]	Financial
Roku TV [9]	Entertainment
Insteon Hub [7]	Smart home
Philips Hue Sync [8]	Entertainment, Smart home
Google Nest [5]	Financial, Entertainment, Smart home

## Chapter 4

### Results

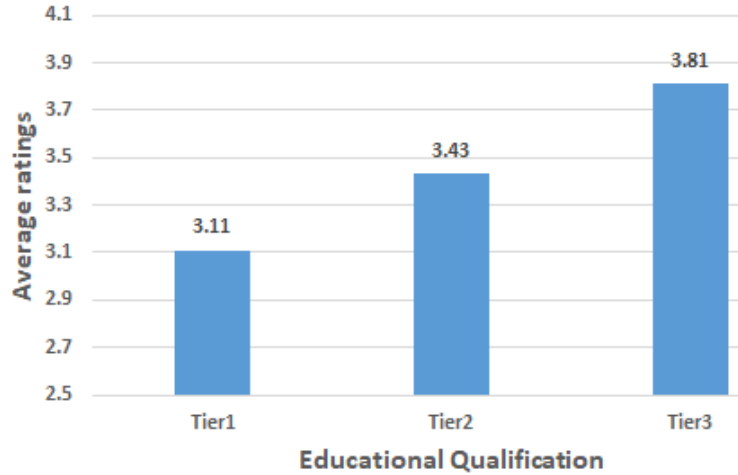
In this section, we explain in detail the results of our survey. Firstly, we discuss the impact of demographic factors on privacy preferences, followed by social media activity factors. Further, we discuss the contextual factors' influence on privacy choices in IOT. To identify all these factors that make a significant impact, we perform an analysis of the ratings given by participants indicating their likeliness in using an application similar to each of the presented 12 IOT scenarios. We consider the two ratings on the likeliness scale, 4, 5 as indicators of liberal mindset regarding privacy concerns whereas the other three ratings- 1,2 and 3 as indicators of conservative thoughts.

Although we cannot directly measure the relationship between privacy concerns for a specific IOT scenario and willingness to use it, we can, however, strongly infer the relationship based on two reasons. Firstly, results from the previous literature indicating the differences in privacy concerns between financial and entertainment contexts confirm our inference [51]. Secondly, the responses from the qualitative survey we conduct regarding privacy concerns strengthen our assumption about this relationship. Moreover, asking the participants about their likeliness to use the application instead of directly probing about their privacy concerns in the survey avoids issues that occur due to privacy paradox. Privacy paradox explains the situation where users even though claim to have concerns regarding the privacy of the information they disclose still tend to go ahead with the disclosure in actuality. [45]

## 4.1 Demographic Factors

We calculate the average likeliness ratings given by participants of all the 12 IOT scenarios shown in the survey. We further analyze the differences in the behavior of participants' distributions derived from their demographic identifiers considered in the study which are Education, Age, Gender, Ethnicity, Continent, Religious views and Marital status. A one-way analysis of variance (ANOVA) with a one-tailed test is used to determine the statistical significance of all such differences in our study. As shown in Fig 4.1, participants in Tier 3 that have either a Master's degree or a Doctorate have comparatively fewer privacy concerns and are more willing to use IOT applications followed by participants in Tier 2 who have finished either a Bachelor's or 4-year college degree. Participants in Tier 1, who have one of the following educational qualifications- Trade/ technical/ vocational training, Associate/ 2-year degree, High School degree or equivalent, Less than a high school diploma have the highest privacy concerns and are least likely in using the applications among all the three educational qualification tiers.

Figure 4.1: Average usage ratings of the participants grouped by Educational Qualifications.

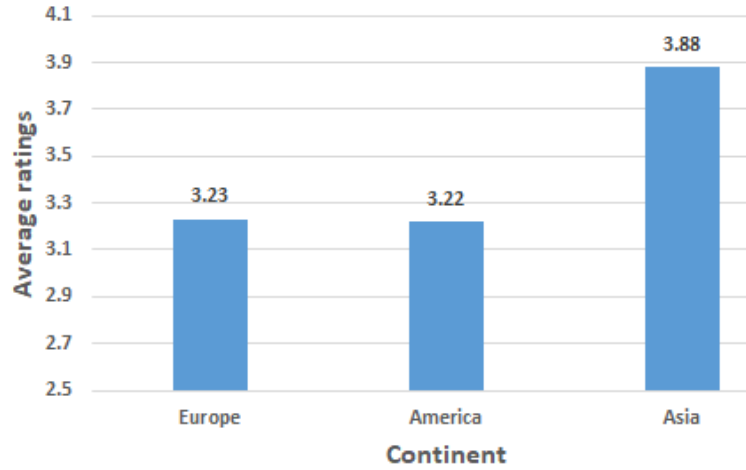


We observe that the higher the educational qualifications acquired by the individuals, the lower the tendency to be bothered regarding IOT privacy. These differences due to educational qualifications are statistically significant ( $p < 0.05$ , F-value=10.8444). We make an interesting



observation on choosing the continent as the demographic identifier and compare the differences in ratings. Participants from Asia as shown in Fig 4.2 have not many privacy concerns in using IOT applications in juxtaposition with participants from the other two continents, Europe and North America. This difference in likeliness ratings between Asia and other continents is statistically significant ( $p < 0.05$ , F-value=17.4232).

Figure 4.2: Average usage ratings of the participants grouped by Continent.

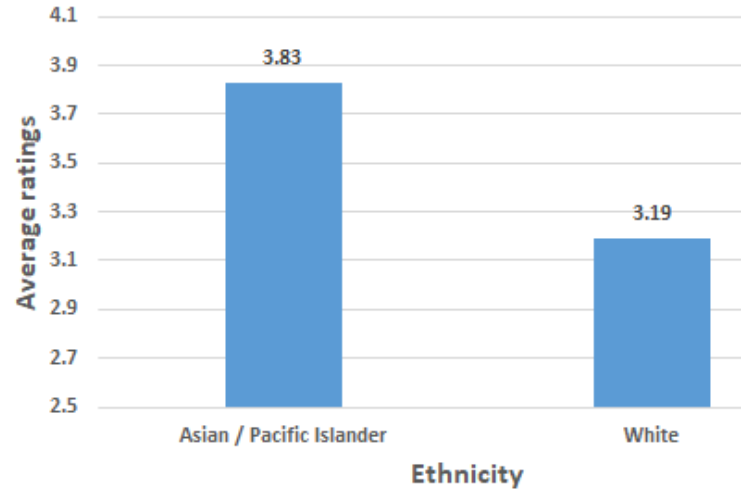


Similar to the earlier observation, as shown in Fig 4.3, we find that participants who identified their ethnicity as Asian/ Pacific Islander are more inclined to use IOT applications and have comparatively fewer privacy issues than the participants who identified their ethnicity as White. This dissimilarity between the ratings with ethnicity as the demographic identifier is statistically significant ( $p < 0.05$ , F-value=16.3978).

Considering marital status as the demographic identifier, as shown in Fig 4.4, we discern that the single participants are comparatively more concerned about their privacy and therefore less likely to use IOT applications than the participants who are married. This variation in participants' ratings of likeliness based on their marital status is statistically significant ( $p < 0.05$ , F-value=27.7543).

Lastly, we notice contrasting privacy-related opinions of participants with respect to the

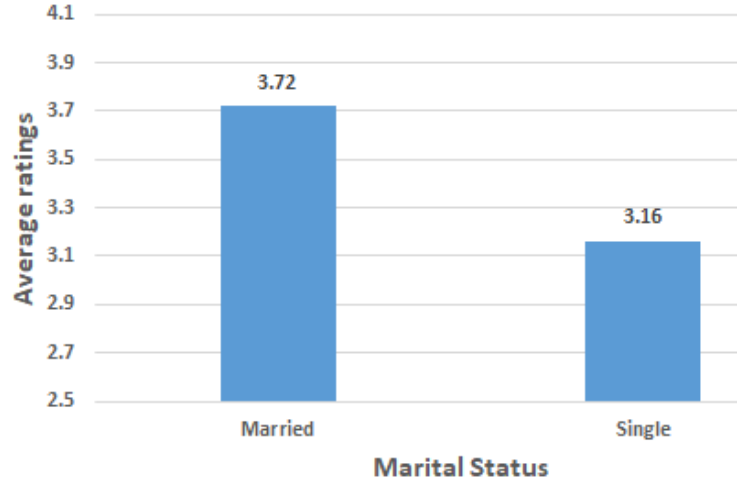
Figure 4.3: Average usage ratings of the participants grouped by Ethnicity.



religious views they follow as the controlling demographic identifier as shown in Fig 4.5. Participants following Hinduism are more liberal-minded about privacy concerns and tend to be highly likely in using IOT applications in comparison to participants either following Christianity or claim to be non-religious. Specifically, non-religious participants are the least probable to use IOT applications owing to their greater privacy concerns. Mainly, we identify that all these distinct behaviors are statistically significant ( $p < 0.05$ , F-value=19.0087).

Although the other two demographic factors, Age and Gender do not have an independent impact on privacy preferences, we find that Age has a statistically significant ( $p < 0.05$ , F-value=5.1681) interaction effect with the factor Continent as shown in Fig 4.6 employing a two-way analysis of variance (ANOVA) and one-tailed test. Participants who reside in North America and are in the age group 18-30 years are more worried about privacy and are less likely to use IOT applications in comparison with those belonging to the age group 31-50 years. Conversely, we find that European participants in the age group 18-30 years are relatively less disturbed with privacy concerns than those in the age group 31-50 years.

Figure 4.4: Average usage ratings of the participants grouped by Marital Status.

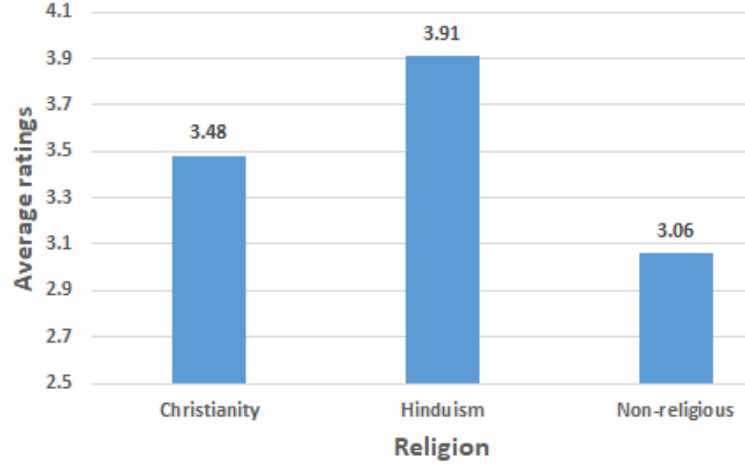


## 4.2 Social Media Activity Factors

We compute a social media activity score for each participant based on their responses describing the number of friends/ followers they have on the social media platform, default privacy settings for posts, time spent daily on social media. In addition to these three identifiers, we also take into consideration participants' agreement ratings about social media being part of their everyday activity and being unhappy with a shutdown of social media. We employ the k-means algorithm for clustering purposes. Based on the results from the elbow method, we make an optimal choice of  $k=4$ . In total, we have 80 participants in Cluster 1, 44 participants in Cluster 2, 53 participants in Cluster 3 and 61 participants in Cluster 4.

Most of the participants in Cluster 1 and Cluster 4 spend at least one hour or more on social media every day. On the contrary, participants in Cluster 2 claim to spend less than two hours and participants in Cluster 3 spend the least amount of time which is less than one hour. Almost all the participants in Cluster 1 and Cluster 4 have at least 200 friends/ followers, whereas, participants in Cluster 3 exhibit exactly opposite behavior and have less than 200 friends/ followers. Cluster 2 represents a nearly equal distribution of participants based on the number of followers/ friends they assert to have on social media.

Figure 4.5: Average usage ratings of the participants grouped by Religion.

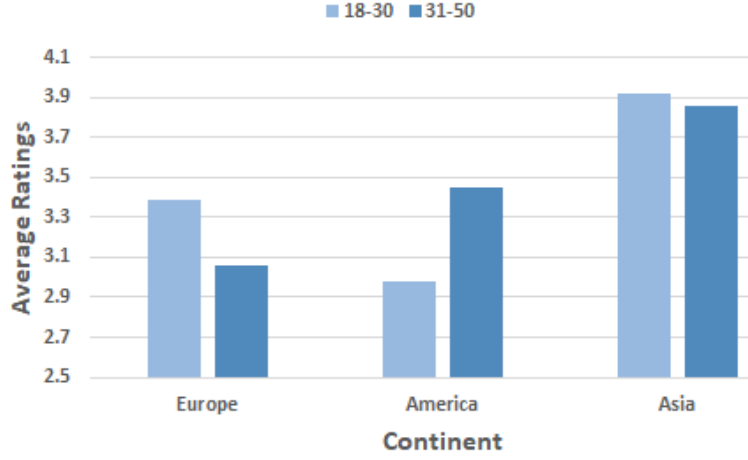


Nearly all participants in the three clusters, Cluster 1, Cluster 3 and Cluster 4, either strongly agree or agree on using social media every day. Specifically, a greater share of participants in Cluster 3 tends to agree rather than strongly agree with the statement. Cluster 2 represents a mixed group of participants with their agreement ratings ranging from strong disagreement to strong agreement. On a detailed look into the ratings for agreement with the assertion of being perturbed due to a complete social media shutdown, we identify that a major part of the participants in Cluster 1, Cluster 3 and Cluster 4 either strongly agree or agree. Moreover, some participants in Cluster 1 and Cluster 4 have a neutral tone regarding social media shutdown. On the other hand participants in Cluster 2 mainly disagree or strongly disagree with this statement.

We find notable distinctions between the default privacy settings for social media posts of participants in each of the four clusters. Participants in Cluster 1 restrict the visibility of their posts to either all of their friends or a few selected friends while participants in Cluster 4 allow their posts to be accessible by either friend of their friends or the whole public. Participants in Cluster 2 let only their friends view their posts whereas Cluster 3 has a motley of participants who permit just friends, friends of their friends and the entire public. We further calculate the average likeliness ratings given by participants in each of the four clusters.

We observe extreme disparities that are statistically significant ( $p < 0.05$ , F-value=9.6641)

Figure 4.6: Average usage ratings of the participants grouped by Age and Continent.

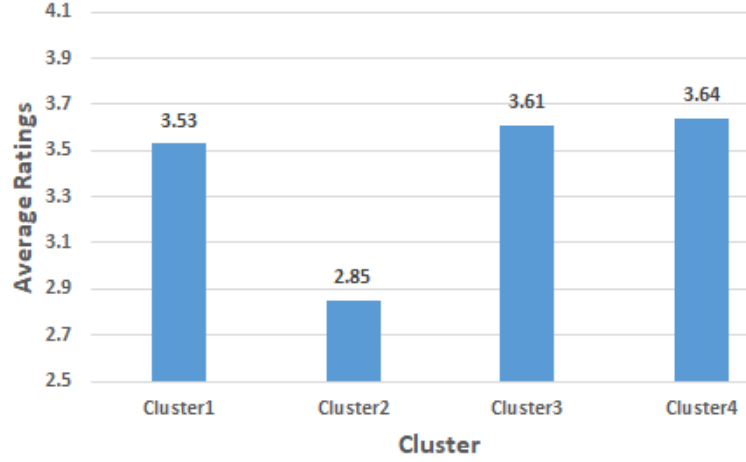


between the ratings specified by participants in Cluster 2 and the remaining three other clusters as shown in Fig 4.7. Participants in Cluster 2 are more concerned with privacy issues and less likely to use IOT applications. This behavior closely resembles their stated social media activity. These participants in comparison to others have fewer followers and spend lesser time on social media. Furthermore, they completely disagree with being affected by any kind of stoppage in social media applications and cautiously carry out their social media operations by setting privacy preferences to only their friends. Overall, relatively less social media activity and restricted privacy choices of these participants are reflected in the ratings given for our survey.

### 4.3 Contextual Factors

We investigate the role played by both the contextual parameters, level of user control on data collection and ultimate utilization purpose on the ratings specified by participants. Initially, we compare the average likeliness ratings given to the four IOT scenarios with a financial use case and the four IOT scenarios with an entertainment use case. The calculated mean of ratings for financial use case scenarios is 3.2 and entertainment use case scenarios is 3.57. This statistically significant ( $p < 0.05$ ,  $F\text{-value}=16.595$ ) difference between the ratings implies that participants are more inclined towards using IOT applications that provide entertainment and are generally

Figure 4.7: Average usage ratings of the participants clustered by Social media activity score.

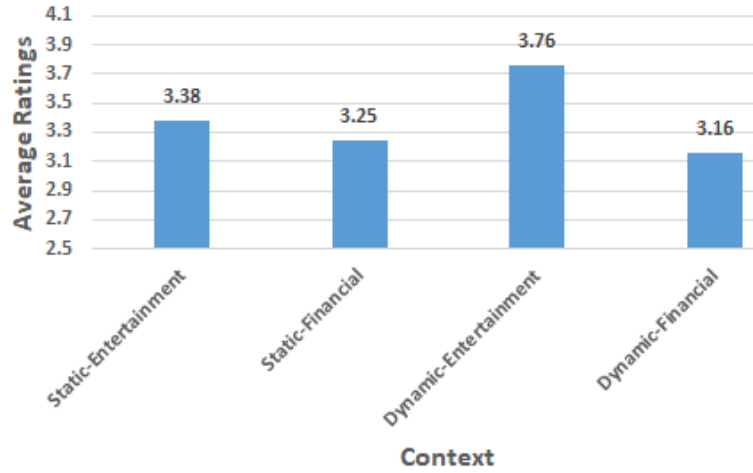


less concerned regarding privacy in such entertainment-related scenarios. On the other hand, participants are more apprehensive about IOT applications used for financial purposes.

Subsequently, we analyze the discrepancies between the average likeliness ratings given to the six IOT scenarios with a static level of user control and the six IOT scenarios with a dynamic level of user control over the application's collection procedure. For static scenarios, the average of ratings is 3.33 and for dynamic scenarios, it is 3.57. This contrast in the ratings is statistically significant ( $p < 0.05$ ,  $F\text{-value}=8.592$ ). An interesting inference to make from this behavior is that participants despite having more control over the information collected in static scenarios prefer to use IOT applications with dynamic user control and have fewer privacy concerns in dynamic scenarios.

Finally, we conduct an exhaustive analysis looking at the differences in ratings between the four distinct combinations of the two contextual parameters in our study - Dynamic- Financial, Dynamic- Entertainment, Static- Financial and Static-Entertainment. We identify a statistically significant difference ( $p < 0.05$ ,  $F\text{-value}=15.282$ ) between ratings for the combinations of static entertainment and dynamic entertainment as shown in Fig 4.8. This insinuates that participants felt more concerned about privacy when they need to provide the information by themselves rather than the entertainment application collecting it in the background. Likewise, we observe a statistically

Figure 4.8: Average usage ratings of the participants based on the scenario's context.



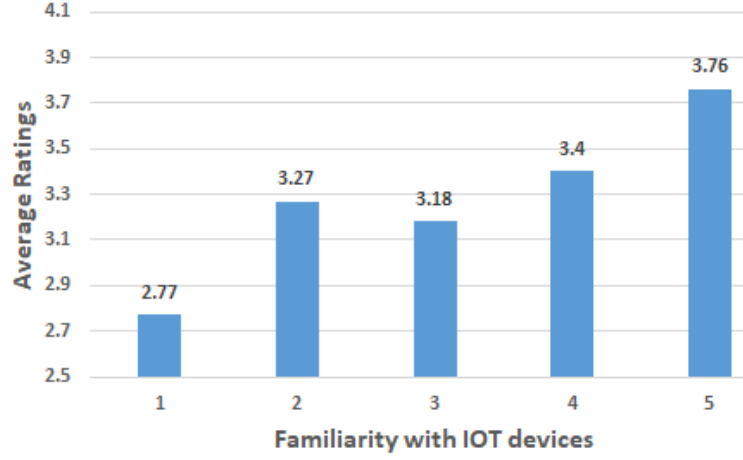
significant difference ( $p < 0.05$ ,  $F\text{-value}=35.397$ ) between ratings for the combinations of dynamic finance and dynamic entertainment. Although participants are less concerned about privacy in dynamic scenarios, they tend to be relatively less willing to use dynamically controlled applications for financial purposes than for entertainment.

#### 4.4 Additional Miscellaneous Factors

Furthermore, we study the impact of miscellaneous factors like familiarity with IOT devices in general, familiarity with each of the given twelve scenarios, possible usage by family/ friends and a million others. In our distribution of participants, we have 16 of them claiming to be least familiar with IOT devices and giving a familiarity rating of one, 14 participants with a familiarity rating of two, 39 participants with a familiarity rating of three, 80 participants with a familiarity rating of four and 89 participants with the highest familiarity rating of five.

We notice that there exists an increasing trend in average likeliness ratings with the increasing IOT familiarity except for the group of participants who have given an IOT familiarity rating of two as shown in Fig 4.9. This statistically significant correlation ( $p < 0.05$  according to spearman's rank-order test results) suggests that participants with a greater degree of familiarity with IOT

Figure 4.9: Average usage ratings of the participants based on their familiarity with IOT devices.



devices are less concerned about any privacy issues with IOT applications. Subsequently, we analyze the influence of the familiarity ratings specifically given by participants for each of the twelve IOT scenarios shown on the corresponding scenario's usage likeliness ratings.

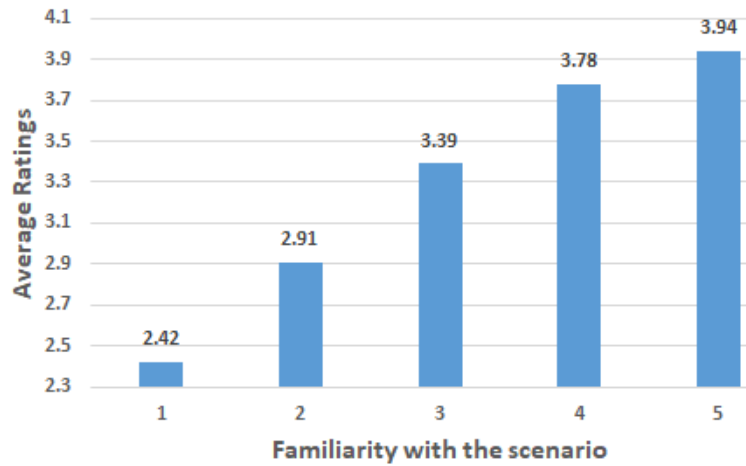
We identify an upward trend in likeliness ratings given by the participants with the increasing familiarity with similar IOT scenarios as shown in Fig 4.10 and this correlation is statistically significant ( $p < 0.05$  according to spearman's rank-order test results). We can infer from this correlation the prominent role of the familiarity with alike IOT scenarios in the willingness to use an IOT application with less concern about privacy. Lastly, we analyze the likeliness ratings given by the participants under the assumptions that their family/ friends or a million others are using it currently. We do not recognize any significant change from the prior behavior of participants when they are asked without these presumptions.

## 4.5 Qualitative Analysis

We conduct qualitative analysis to understand the primary concerns that the participants have regarding different IOT application scenarios in general. We provide the participants with our definitions of static and dynamic levels of user control over the data collection by IOT applications and then gather their opinions regarding privacy concerns in the four different IOT scenario



Figure 4.10: Average usage ratings of the participants based on their familiarity with the given IOT scenario.



categories which are Dynamic- Financial, Dynamic- Entertainment, Static- Financial, Static- Entertainment. Firstly, we examine the participants' responses for privacy concerns in the hypothetical IOT scenario with static control used for entertainment purposes and categorize them mainly into three types as shown in Table 4.1.

The first category, hacking or compromise of the data shared with the IOT app is the chief concern of 78 participants. For 39 other participants in the second category, the possibility of their data being either sold to or shared with third parties for commercial purposes causes apprehension about their privacy. The requirement to share sensitive information like fingerprints or passwords is the reason for worry about privacy for 36 participants in the third category. Responses in the first two categories elucidate the concerns participants have about the aftermath of sharing data with the IOT apps, whereas, responses in the third category explain the inconvenience faced by participants in sharing the data itself due to its high perceived sensitivity. We then calculate the average of likeliness ratings given to the two Static- Entertainment scenarios by these three categories of participants separately.

Participants whose responses are in the second category are relatively less likely to use these applications with an average rating value of 2.76, while participants with comments in the third

Table 4.1: Privacy Concerns of participants in Static Entertainment IOT scenarios

Category	# Turkers	Sample Responses
Hacking, data compromise and identity theft	78	<ol style="list-style-type: none"> <li>1. I would be concerned about data breaches in which my data is compromised.</li> <li>2. That they can be used to steal and use my identity illegally.</li> </ol>
Sold to or shared with third parties for commercial purposes	39	<ol style="list-style-type: none"> <li>1. I would be afraid that my personal data could be used by third parties for commercial purposes.</li> <li>2. I would be concerned that such information might be sold to a third party for advertisement purposes and it could result in annoying advertisements.</li> </ol>
Concerns regarding sensitive information like fingerprints, passwords	36	<ol style="list-style-type: none"> <li>1. Static data like fingerprint, passwords is quite sensitive and I would not be happy providing it.</li> <li>2. The static information about age, gender its okay but I have concerns about fingerprints.</li> </ol>

category are more likely to utilize such applications with a mean rating of 3.49. These participants are followed by those having their concerns in the first category with a rating of 3.31. Secondly, we investigate the responses given by participants when quizzed about privacy concerns associated with the IOT scenario in which the information is collected dynamically by the entertainment providing app as shown in Table 4.2

Similar to the responses received for the previous IOT scenario, Static- Entertainment, 35 participants in the first category cited hacking or compromise of their information communicated with the IOT application as the reason for privacy concern in Dynamic- Entertainment scenarios. Likewise, the possible incurrence of monetary benefits by the application service providers with the information they shared seems like a privacy threat to 28 participants in the second category. Constant tracking of activity or location seems to be causing inconvenience to the privacy of 79 participants in the third category. We notice that the major difference in participants' responses between static and dynamic scenarios of IOT applications used for entertainment purposes is the discussion about an inkling of being monitored all the time.

On the further computation of the average of likeliness ratings given to the two Dynamic- Entertainment scenarios by these three categories of participants, we find that participants with

responses in the first and third categories are comparatively less concerned about privacy in these scenarios with mean rating values of 3.7 and 3.72 respectively. Participants having their responses in the second category tend to be less likely to use these kinds of IOT applications and their average likeliness rating is 3.34. Next, we inspect the responses regarding participants' privacy concerns in Static- Financial scenarios and classify them into two categories as shown in Table 4.3.

The first category, hacking or compromise of the exchanged information seems to be a recurring reason for privacy concerns and this has been mentioned in the responses for Static- Financial IOT scenarios by 94 participants. Furthermore, 40 participants in the second category deemed certain financial information like passwords, banking details to be highly sensitive and are uncomfortable in sharing it with IOT apps. From these responses, we can infer that when it comes to IOT scenarios with static control over information collected by the applications, irrespective of the motivation in using these applications being entertainment or financial, participants take into consideration for privacy the sensitivity of shared information.

Following this, we evaluate the average of likeliness ratings given to the two Static- Financial scenarios by these two categories of participants. We observe that participants with responses in the first and second categories have almost similar mean rating values of 2.98 and 3.09 respectively. Finally, we look into the responses about privacy concerns in Dynamic- Financial scenarios and group them into three categories as shown in Table 4.4. Hacking or data compromise by the financial IOT applications is the paramount privacy concern for 59 participants in the first category.

Commercial gains achievable by IOT application service providers on sharing or selling user data to third parties is considered as the cause for privacy concern in 44 participants in the second category. Concerns have been raised about the continuous financial activity tracking by 46 participants in the third category. Regardless of the utilization of the IOT application for entertainment or financial purposes, we can conclude from these responses that with regards to IOT scenarios collecting information dynamically in the background, privacy concerns will be raised about constant activity tracking.

Calculation results of the average of likeliness ratings given to the two Dynamic- Financial

scenarios by these three categories of participants indicate that participants with responses in the second category are in comparison to the other two categories less likely to use these IOT applications with mean rating value of 2.65. Participants having responses in the third category tend to be less concerned about Dynamic- Financial scenarios with an average rating value of 3.04 followed by those with responses in the first category who have a mean value of 2.79. Overall, participants show concerns about safeguarding their privacy in the event of a malicious attack like hacking in all the four IOT scenarios they are presented with. Besides this, we also observe that being inundated with advertisements or recommendations from third parties based on the information participants shared with IOT application service providers is one other common privacy concern across the IOT scenarios.

#### 4.6 Privacy Prediction

We have successfully answered the four research questions- RQ1, RQ2, RQ3 and RQ4 with the previously discussed results. In this section, we employ these findings of the significant factors influencing privacy choices to design a privacy preference classification model. We identify that the demographic factors Continent, Educational Qualifications, Ethnicity, Marital status and Religious beliefs have a prominent role in making privacy choices about IOT applications. In addition to these factors, Age also has an important part in privacy decisions when considered together with the Continent of the individual.

All these demographic factors are selected as features in training our classification models. We then include the social media activity score computed earlier as another feature to represent the impact of all the social media factors collectively on privacy concerns. Another significant role-playing factor, familiarity rating with IOT devices is also chosen as one of the features. Finally, we consider the combination of the contextual factors, category of IOT scenario (Dynamic- Financial, Dynamic- Entertainment, Static- Financial, Static- Entertainment) as a feature in addition to the familiarity rating with the corresponding scenario.

For the purpose of classification, we define a dataset by separating each participant's responses

for different IOT scenario categories into eight individual tuples. Thus in total, we have 1920 tuples in our dataset. As mentioned earlier, we classify the privacy preferences of participants into two categories Liberal and Conservative based on the likeliness rating values. Hence, our main goal is to build a machine learning-based binary classification model to predict participant’s likeliness to use IOT applications similar to the discussed IOT scenarios. On a whole, we have 1056 tuples with privacy classifier labels as liberal and 848 tuples as conservative. We split this dataset further into training and testing datasets in a 70:30 ratio.

The training dataset has 721 tuples with the liberal category for privacy preferences and 611 tuples with the conservative category. To handle this existing imbalance in the tuple count of classifier categories of the dataset and thereby reduce misclassification errors, we utilize an approach called SMOTE (Synthetic Minority Oversampling Technique) [22]. After the successful application of the SMOTE approach, by synthetically creating minority classifier category tuples in order to oversample minor category, we have in the training dataset, 721 tuples each for both the privacy classifier labels, Liberal and Conservative.

#### 4.6.1 XGBoost Classifier

Initially, we use XGBoost (Extreme Gradient Boosting) algorithm to design the binary classifier. We make this choice due to XGBoost’s high scalability in a wide range of scenarios and efficiency in classification problems [23]. This bagging-based decision tree algorithm takes a parallelized, distributed approach for quicker learning rates and uses gradient descent for boosting to minimize the error rates. We perform hyperparameter tuning using a Python library, Hyperopt [19] to adjust hyperparameters for an optimized performance of the underlying machine learning algorithm.

Based on these tuning results, we identify the following best hyperparameters for privacy classification XGBoost model - maximum depth of the tree (max\_depth) needs to be 14, learning\_rate of 0.26, minimum loss reduction (gamma) of value zero, the minimum sum of instance weight required in child node (min\_child\_weight) be value eight, a 0.96 subsample ratio of training

instances (subsample) and a 0.74 subsample ratio of columns in each tree (colsample\_bytree). Using these hyperparameters we train our classification model and make privacy predictions. We achieve an accuracy of 71.64% and a Cohen’s kappa score of 0.4328 in predicting classifier labels indicating an individual’s privacy concern levels.

All the features used in training this XGBoost classification model are scored based on the total number of times the feature has been used as a decision-maker to split a node of a decision tree. The rating value of familiarity with a given IOT scenario is the most important feature in prediction classification with a score of 0.27. This feature is followed by Continent and Ethnicity features with importance scores of 0.13 and 0.12 respectively. Familiarity with IOT devices, Marital status and Religion features have their importances scores ranging between 0.07 and 0.09. Three other features Educational qualifications, Social media activity score and Age have a score of about 0.06. Out of all the features, the IOT scenario category has the least importance score of 0.05.

Table 4.2: Privacy Concerns of participants in Dynamic Entertainment IOT scenarios

Category	# Turkers	Sample Responses
Hacking, data compromise	35	<ol style="list-style-type: none"> <li>1. I would be concerned that such information might eventually be released through means of a data breach.</li> <li>2. That my dynamic info could be leaked and used against me in public shaming.</li> </ol>
Sold to or shared with third parties for commercial purposes	28	<ol style="list-style-type: none"> <li>1.No real privacy concerns about dynamic info other than my data being monetised by a third party.</li> <li>2.Getting flooded with "personalized" ads from that app or "third parties" the app sells their data to.</li> </ol>
Concerns regarding continuous monitoring and type of information collected like location, activity tracking	79	<ol style="list-style-type: none"> <li>1. That would be slightly more annoying, I wouldn't be able to shake the feeling that I'm constantly "under surveillance".</li> <li>2. I am more apprehensive in sharing my location, activities and preferences as it can be crucial for my safety.</li> </ol>

#### 4.6.2 SVM Classifier

Subsequently, we also build a binary classification model using another popular algorithm called SVM (Support Vector Machine). Selection of this algorithm is done considering its cost-efficient classification techniques to build an optimal hyperplane for differentiating different categories with a higher dimensional dataset [31]. We specifically use the SVM classifier provided by the Python library scikit-learn [49]. We also use GridSearchCV from the same library for hyperparameter tuning purposes. We obtain the following best-fit hyperparameters for our SVM classifier-regularization parameter (C) with a value 10 and a linear kernel type.

Using these hyperparameter settings, we get an accuracy of 72.09% and a Cohen's kappa score of 0.4418. Features involved in this algorithm are given scores of importance based on the weights assigned to them. Familiarity with the IOT scenario in consideration continues to be the most important factor in this classifier algorithm also with a score of 0.59. Importance ranking is then followed by two features Ethnicity and Educational Qualifications with corresponding scores of 0.27 and 0.24. The respective scores of the features Religion and IOT scenario category are 0.16 and 0.14. The four features Age, Marital status, Social media activity score have scores ranging from 0.05 to 0.1. Familiarity with IOT devices is the least significant feature with a score of 0.02. Therefore, out of the chosen two classifiers, the SVM classifier performs slightly better based on the metrics, accuracy and Cohen's kappa score.

### 4.7 Privacy Awareness Assistant

We propose a personalized privacy awareness assistant applying the binary privacy classification model we build earlier. In this web-based application built with Flask framework, we use HTML5, CSS for designing the user interface and MongoDB as the database to store annotated privacy policies. After a thorough examination of privacy policies of the selected 13 IOT applications, we identify the following concerns that are being commonly addressed across applications:

- **Concern 1:** Do users have the ability in the future to change privacy preferences?

Table 4.3: Privacy Concerns of participants in Static Financial IOT scenarios

Category	# Turkers	Sample Responses
Hacking, data compromise	94	<ol style="list-style-type: none"> <li>1. They should keep it safe, encrypted and not share it with anyone else.</li> <li>2. Losing money because of someone making purchases via my account if it gets hacked; or if the whole app' database gets hacked.</li> </ol>
Uncomfortable about sharing sensitive financial information like passwords, bank details	40	<ol style="list-style-type: none"> <li>1. Static data for financial purposes is a major concern of worry for me since that would include my bank and credit card details, which are too sensitive.</li> <li>2. I would not want the app to collect my credit card details, password, fingerprints, otherwise ok.</li> </ol>

- **Concern 2:** How easy is it for the users to change privacy preferences if they can?
- **Concern 3:** Should the user set privacy preferences once initially or update regularly based on the newly collected data?
- **Concern 4:** Is consent required when information is shared with third parties?
- **Concern 5:** Do the users have a provision to opt-out of personalization of service?
- **Concern 6:** Do the users have a provision to opt-out of receiving targeted advertisements?
- **Concern 7:** Can the users delete or modify collected information?
- **Concern 8:** Can users verify if the information collected is accurate?
- **Concern 9:** Is the information shared with third parties anonymized?
- **Concern 10:** What is the followed data retention policy?
- **Concern 11:** What security practices are followed to safeguard your data?
- **Concern 12:** How policy changes will be conveyed to the user?
- **Concern 13:** Will any warnings be sent to the user in case of a privacy breach?



- **Concern 14:** If the application delivers a continuous monitoring service, does it record all the information or can you control it?

Figure 4.11: Questionnaire about demographic information in Privacy Awareness Assistant

What age group do you fall in? 18-30 yrs ▼

What is your Educational Background? Trade/technical/vocational training ▼

Which continent do you currently reside in? North America ▼

What best describes your race/ethnicity? Asian / Pacific Islander ▼

What religious view do you follow? Christianity ▼

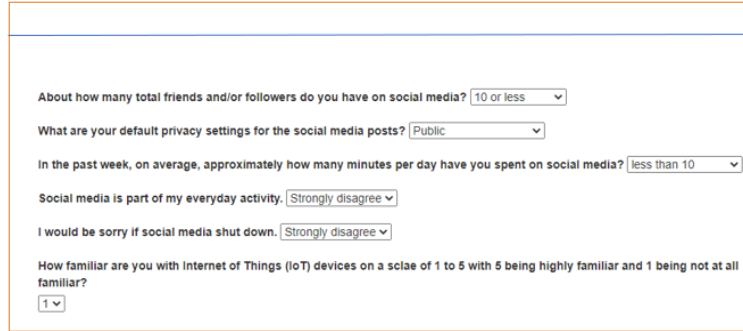
What is your marital status? Married ▼

Additionally, in the above-mentioned list, we attempt to include the primary privacy concerns expressed by our qualitative survey participants about the possible data breach incidents, third party sharing and continuous monitoring. Information about Concern 11 and Concern 13 can help in addressing user's concerns about data hacking or compromise situations in general. Answers to Concern 4, Concern 6 and Concern 9 can collectively address concerns about information sharing with third parties. Also, concern about continuous monitoring services can be answered with Concern 14. For all the privacy policies that our privacy assistant processes as part of user queries, annotation takes place to fetch information regarding these 14 concerns and any additional concerns user has. A typical functional workflow in our web application proposal is as follows:

**Step 1:** User provides information about their Age, Continent they currently reside in, Ethnicity/ Race, Religious views they follow, Educational background and Marital status as shown in Fig 4.11.

**Step 2:** User, then gives information about their social media activity like the number of followers or friends, time spent on social media in a week, default privacy settings for posts on social media as shown in Fig 4.12.

Figure 4.12: Questionnaire about social media activity in Privacy Awareness Assistant



About how many total friends and/or followers do you have on social media?

What are your default privacy settings for the social media posts?

In the past week, on average, approximately how many minutes per day have you spent on social media?

Social media is part of my everyday activity.

I would be sorry if social media shut down.

How familiar are you with Internet of Things (IoT) devices on a scale of 1 to 5 with 5 being highly familiar and 1 being not at all familiar?

**Step 3:** Lastly, the application collects information about the IOT application the user wants to determine his privacy preferences for as shown in Fig 4.13

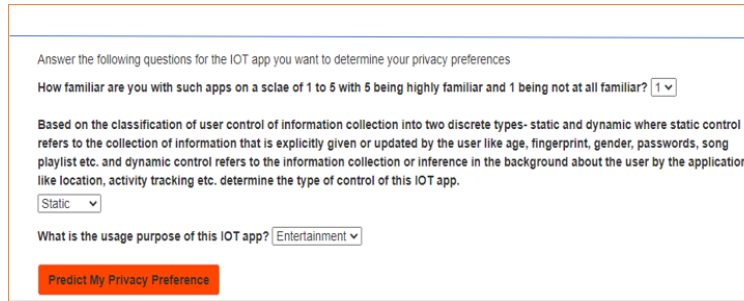
**Step 4:** Then the web application processes this user-given information with the help of the binary classifier model we discussed earlier to predict the user's privacy choice either Liberal or Conservative in using the IOT application.

**Step 5:** In addition to learning their predicted privacy choice, users can query about specific privacy concerns they have as shown in Fig 4.14. Users have an option to choose from a list of IOT applications that have already been annotated by our tool. In case, our web application never processed before the privacy policy of that IOT application users should upload the corresponding privacy policy for the IOT application. Also, in our endeavor to give maximum assistance to inexperienced users, we provide a list of 14 concerns they can pick from. If the users have other particular concerns they can submit their query for it instead.

**Step 6:** Finally, the web application displays the user's query results after annotating the privacy policy. For example, if the IOT application, user queries for is Alexa and the concern is chosen from the list which is- *Is consent required when information is shared with third parties?* then the result will be as follows:

*There are also thousands of skills created by third-party developers, everything from quizzes to games to meditations, which you can access through Alexa. When you use a third-party skill, we will exchange related information with that third party so they can respond to your request – for instance,*

Figure 4.13: Questionnaire about IOT application in Privacy Awareness Assistant



Answer the following questions for the IOT app you want to determine your privacy preferences

How familiar are you with such apps on a scale of 1 to 5 with 5 being highly familiar and 1 being not at all familiar?

Based on the classification of user control of information collection into two discrete types- static and dynamic where static control refers to the collection of information that is explicitly given or updated by the user like age, fingerprint, gender, passwords, song playlist etc. and dynamic control refers to the information collection or inference in the background about the user by the application like location, activity tracking etc. determine the type of control of this IOT app.

What is the usage purpose of this IOT app?

*your answers when you play a trivia skill. We don't share any personally identifiable information with that third party without your agreement. For example, a restaurant booking skill might ask for your email address to send you confirmation of your reservation. We would only share that information with your permission. You can see and manage the skills that have requested permission to access data here. None of your voice recordings are ever shared with third-party skill developers.*

The web application stores only the IOT application-related information about privacy concerns and policies but not any other user-related information. To ensure secure communication between the user and the web application, the application uses HTTPS. Also, the application sanitizes the entered queries to prevent any Cross-Site Scripting (XSS) attacks.

Figure 4.14: Questionnaire about privacy concern in Privacy Awareness Assistant

We predict that you will be **Conservative** with your privacy choices in using this IOT app

If you have any specific privacy concerns about this app you want to be informed about, answer the following questions.

What is the name of this IOT app?

If chosen other previously name this app.

If chosen other previously enter the provided privacy policy of this app.

What is your main privacy concern in using this IOT app?

If chosen other previously enter your privacy concern here.

Table 4.4: Privacy Concerns of participants in Dynamic Financial IOT scenarios

Category	# Turkers	Sample Responses
Hacking, data compromise	59	<ol style="list-style-type: none"> <li>1. I would be worried that someone might be able to breach into the application and cause significant financial harm to me.</li> <li>2. If the collection of data is well protected, I would have no problem.</li> </ol>
Sold to or shared with third parties for commercial purposes	44	<ol style="list-style-type: none"> <li>1. Third-party companies could use the dynamic information to spam email or text alerts if I'm near their locations.</li> <li>2. They might sell my information to credit or insurance companies that might try to sell me their products/ plans.</li> </ol>
Concerns regarding activity tracking mainly financial transactions	46	<ol style="list-style-type: none"> <li>1. I would be concerned that they would be using it to track the location of ATMs and potential credit card use areas.</li> <li>2. I would feel like it was an invasion of privacy because someone would know how and what I spend my money on.</li> </ol>

## Chapter 5

### Discussion

Earlier work suggests that users prefer to ignore the underlying privacy risks when sharing information with IOT applications for a more personalized service [32]. This tradeoff behavior encourages IOT service providers to focus primarily on customizing the application's performance suiting the user's personality. But this compromise between providers and users is a constant looming privacy concern. In the past years, several security breach incidents surfaced that endangered the privacy of IOT application users like the misuse of shared information, malicious malware attacks, ineffective user data encryption [56], [62]. Such events might eventually reduce the confidence of users in IOT applications and prevent them from utilizing these applications.

To avoid all these detrimental consequences, safeguarding the privacy of users should be considered as an utmost important task by the service providers. However, privacy preservation presents a huge challenge to the service providers because a wide range of factors like user demographics, context, data collection control can impact privacy preference settings which need to be taken into account. The findings in our work help service providers understand the significant role played by user's social media activity and certain demographic factors like Continent, Educational Qualifications, Ethnicity, Marital status and Religious beliefs in determining privacy preferences.

The majority of current IOT applications continuously monitor the users and collect data in the background to deliver a better-customized service. Although this decreases the degree of control users can have on the data collection, our participants exhibited more willingness and expressed fewer privacy concerns in using such applications rather than those where they are required to

provide or update information explicitly. This observation highlights the hesitation faced by IOT users in assuming complete control over data sharing. Furthermore, prior studies reveal that users are relatively more cautious about sharing information with IOT applications for financial purposes [51]. Our survey responses echo these findings and show that irrespective of the level of control assumed, users, are less comfortable with financial IOT applications unlike IOT applications providing entertainment.

Applying all findings from our survey, we successfully build a privacy preference classification model. This demonstrates the feasibility to predict the privacy preferences of users based on demographic, social media activity and contextual factors. Our research provides directions for future work in the implementation of privacy-preserved IOT applications. IOT application service providers can design different versions to cater to the various categories of users based on their predicted privacy preferences. Moreover, it is important to develop tools to empower users to overcome their hesitation in facing privacy challenges and assume collective responsibility with service providers in safeguarding their privacy. In this regard, we come up with the design of a privacy awareness assistant which is a web application that can guide users to make informed decisions about their privacy. This assistant can also result in removing any unwarranted privacy concerns that the inexperienced users might have about an IOT application by conveying information provided by the privacy policy in a concise manner.

## 5.1 Limitations

We acknowledge certain limitations in our current study procedure that need to be addressed in the future. Firstly, our study is conducted on Amazon Mechanical Turk where on a whole 240 participants are surveyed. Although this is comparatively a smaller representation, all the results presented in this work are statistically significant which helped us in achieving the primary purpose of answering the four research questions RQ1, RQ2, RQ3 and RQ4 about the factors influencing privacy preferences in IOT applications. Besides, we ensure an equal representation of participants in the continent, gender, and age demographics. Moreover, prior research propounds

on the reliability of Amazon Mechanical Turkers responses in developing privacy setting interfaces in IOT applications [17]. Second, our survey setting where participants are required to give likeliness ratings in using IOT applications could have prompted them to think in a more analytical manner unlike in the real world. However, collecting likeliness ratings instead of directly asking participants to provide their privacy preferences for the IOT application scenarios helped in avoiding the possible situation of privacy paradox.

Third, sample IOT scenario situations chosen in our study are limited by variations in two contextual parameters, level of user control over information collection and utilization purpose. In the future, we plan to include another contextual parameter, incorporated mechanism for data security and study its influence in conjunction with other parameters on privacy preferences in IOT applications. This particular choice of parameter is made based on the analysis of responses to our qualitative survey where most of the participants expressed concerns like hacking or data compromise regarding the security of the collected data. Fourth, the majority of our participants are familiar with IOT devices in general and one reason for this could be the fact that this survey is conducted on an online platform where most workers are active internet users. To account for the privacy concerns of non-IOT familiar people, in the future we will reach out to such people and conduct in-person interviews.

Fifth, the privacy preference classification model we designed can be prone to unidentified bias. A prior study explains the inadvertent effects of bias in data and algorithm on the fairness of the machine learning model’s predictions [41]. In this regard, we propose to mitigate any such concerns in our subsequent work and ensure a fair model for determining privacy preferences. Finally, our privacy awareness assistant is a development in progress. Our future work will mainly focus on adopting the existing annotation tools in the market like TagTog [21], WebAnno [60] and Prodigy into our architecture which can actively learn from the feedback on predicted annotations. To entirely automate the annotation process, the selected tool needs to train with hundreds of privacy policies for accurate results. Thus, creating such a dataset will be our subsequent task.

## **Chapter 6**

### **Conclusion**

With the intent of assisting users in resolving their dilemma about privacy in using IOT applications, we have conducted a multi-continental survey and identified the significant impact of certain demographic, social media activity and contextual factors on privacy preferences. Mainly demographic factors- Education, Ethnicity, Continent, Religion, Marital status play a crucial role in privacy choice making. Our study sheds light on the hesitation faced by individuals when required to assume complete control over data provision and updation rather than allowing IOT applications to perform continuous data collection in the background. Furthermore, we find that individuals tend to be more concerned about privacy in using financial applications unlike entertainment applications regardless of the level of control on data collection. All these findings have paved the way for the building of a personalized privacy prediction classifier with 72% accuracy and thereby we have managed to propose the design of a privacy awareness assistant that can guide the users in the privacy decision-making process.



## Bibliography

- [1] Alexa Privacy Policy,. <https://www.alexa.com/help/privacy>.
- [2] Apple TV Privacy Policy,. <https://support.apple.com/en-us/HT208511>.
- [3] Fitbit Privacy Policy,. <https://www.fitbit.com/global/us/legal/privacy-policy>.
- [4] Garmin Pay Privacy Policy,. <https://www.garmin.com/en-US/privacy/garminpay/policy/>.
- [5] Google Nest Privacy Policy,. <https://nest.com/legal/privacy-statement-for-nest-products-and-services/>.
- [6] Honeywell Total Connect Comfort Privacy Policy,. <https://mytotalconnectcomfort.com/portal/Account/TermsAndConditions>.
- [7] Insteon Hub Privacy Policy,. <https://www.insteon.com/privacy>.
- [8] Philips Hue Sync Privacy Policy,. <https://www.philips-hue.com/en-us/support/legal/privacy-policy>.
- [9] Roku Privacy Policy,. <https://docs.roku.com/published/userprivacypolicy/en/us>.
- [10] Samsung Family Hub Refrigerator Privacy Policy,. <https://account.samsung.com/membership/etc/specialTC.do?fileName=familyhub2.html>.
- [11] SimpliSafe Privacy Policy,. <https://simplisafe.com/privacy>.
- [12] Wemo WiFi Smart Plug Privacy Policy,. <https://www.wemo.com/privacy-notice/>.
- [13] Xfinity Voice Remote Privacy Policy,. <https://www.xfinity.com/privacy>.
- [14] Worldwide Internet of Things Forecast,. [https://www.idc.com/getdoc.jsp?containerId=IDC\\_P24793](https://www.idc.com/getdoc.jsp?containerId=IDC_P24793), 2017.
- [15] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. Privacy and human behavior in the age of information. *Science*, 347(6221):509–514, 2015.
- [16] Ivor D Addo, Sheikh I Ahamed, Stephen S Yau, and Arun Buduru. A reference architecture for improving security and privacy in internet of things applications. In 2014 IEEE International conference on mobile services, pages 108–115. IEEE, 2014.

- [17] Paritosh Bahirat, Yangyang He, Abhilash Menon, and Bart Knijnenburg. A data-driven approach to developing iot privacy-setting interfaces. In 23rd International Conference on Intelligent User Interfaces, pages 165–176, 2018.
- [18] Steven Bellman, Eric J Johnson, Stephen J Kobrin, and Gerald L Lohse. International differences in information privacy concerns: A global survey of consumers. The Information Society, 20(5):313–324, 2004.
- [19] James Bergstra, Dan Yamins, and David D Cox. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In Proceedings of the 12th Python in science conference, volume 13, page 20. Citeseer, 2013.
- [20] Nellie Bowles. Thermostats, locks and lights: Digital tools of domestic abuse. The New York Times, 23, 2018.
- [21] Juan Miguel Cejuela, Peter McQuilton, Laura Ponting, Steven J Marygold, Raymund Stefancsik, Gillian H Millburn, Burkhard Rost, FlyBase Consortium, et al. tagtog: interactive and text-mining-assisted annotation of gene mentions in plos full-text articles. Database, 2014, 2014.
- [22] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16:321–357, 2002.
- [23] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794, 2016.
- [24] Hichang Cho, Milagros Rivera-Sánchez, and Sun Sun Lim. A multinational study on online privacy: global concerns and local responses. New media & society, 11(3):395–416, 2009.
- [25] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, and Julie A Kientz. Living in a glass house: a survey of private moments in the home. In Proceedings of the 13th international conference on Ubiquitous computing, pages 41–44, 2011.
- [26] Lorrie Cranor. Web privacy with P3P. ” O’Reilly Media, Inc.”, 2002.
- [27] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. IEEE Pervasive Computing, 17(3):35–46, 2018.
- [28] Cailing Dong, Hongxia Jin, and Bart P Knijnenburg. Ppm: A privacy prediction model for online social networks. In International Conference on Social Informatics, pages 400–420. Springer, 2016.
- [29] Lujun Fang and Kristen LeFevre. Privacy wizards for social networking sites. In Proceedings of the 19th international conference on World wide web, pages 351–360, 2010.
- [30] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. How short is too short? implications of length and framing on the effectiveness of privacy notices. In Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016), pages 321–340, 2016.

- [31] Sasan Karamizadeh, Shahidan M Abdullah, Mehran Halimi, Jafar Shayan, and Mohammad javad Rajabi. Advantage and drawback of support vector machine functionality. In 2014 International conference on computer, communications, and control technology (I4CT), pages 63–65. IEEE, 2014.
- [32] Dongyeon Kim, Kyuhong Park, Yongjin Park, and Jae-Hyeon Ahn. Willingness to provide personal information: Perspective of privacy calculus in iot services. Computers in Human Behavior, 92:273–281, 2019.
- [33] Jan Kolter, Thomas Kernchen, and Günther Pernul. Collaborative privacy management. computers & security, 29(5):580–591, 2010.
- [34] K Krishna and M Narasimha Murty. Genetic k-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(3):433–439, 1999.
- [35] Marc Langheinrich. A privacy awareness system for ubiquitous computing environments. In international conference on Ubiquitous Computing, pages 237–245. Springer, 2002.
- [36] Hosub Lee, Richard Chow, Mohammad R Haghighat, Heather M Patterson, and Alfred Kobsa. Iot service store: A web-based system for privacy-aware iot service discovery and interaction. In 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pages 107–112. IEEE, 2018.
- [37] Hosub Lee and Alfred Kobsa. Privacy preference modeling and prediction in a simulated campuswide iot environment. In 2017 IEEE International Conference on Pervasive Computing and Communications (PerCom), pages 276–285. IEEE, 2017.
- [38] Hosub Lee and Alfred Kobsa. Towards ubiquitous privacy decision support: Machine prediction of privacy decisions in iot. In Convergence of Artificial Intelligence and the Internet of Things, pages 87–115. Springer, 2020.
- [39] In Lee and Kyoochun Lee. The internet of things (iot): Applications, investments, and challenges for enterprises. Business Horizons, 58(4):431–440, 2015.
- [40] Yao Li, Alfred Kobsa, Bart P Knijnenburg, and MH Carolyn Nguyen. Cross-cultural privacy prediction. Proceedings on Privacy Enhancing Technologies, 2017(2):113–132, 2017.
- [41] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635, 2019.
- [42] Sharad Mehrotra, Alfred Kobsa, Nalini Venkatasubramanian, and Siva Raj Rajagopalan. Tip-pers: A privacy cognizant iot environment. In 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops), pages 1–6. IEEE, 2016.
- [43] Sandra J Milberg, Sandra J Burke, H Jeff Smith, and Ernest A Kallman. Values, personal information privacy, and regulatory approaches. Communications of the ACM, 38(12):65–74, 1995.
- [44] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017), pages 399–412, 2017.

- [45] Patricia A Norberg, Daniel R Horne, and David A Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of consumer affairs, 41(1):100–126, 2007.
- [46] Dara O’Neil. Analysis of internet users’ level of online privacy concerns. Social Science Computer Review, 19(1):17–31, 2001.
- [47] Primal Pappachan, Martin Degeling, Roberto Yus, Anupam Das, Sruti Bhagavatula, William Melicher, Pardis Emami Naeini, Shikun Zhang, Lujo Bauer, Alfred Kobsa, et al. Towards privacy-aware smart buildings: Capturing, communicating, and enforcing privacy policies and preferences. In 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), pages 193–198. IEEE, 2017.
- [48] Niklas Paul, Welderufael B Tesfay, Dennis-Kenji Kipker, Mattea Stelter, and Sebastian Pape. Assessing privacy policies of internet of things services. In IFIP International Conference on ICT Systems Security and Privacy Protection, pages 156–169. Springer, 2018.
- [49] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. the Journal of machine Learning research, 12:2825–2830, 2011.
- [50] Pawani Porambage, Mika Ylianttila, Corinna Schmitt, Pardeep Kumar, Andrei Gurtov, and Athanasios V Vasilakos. The quest for privacy in the internet of things. IEEE Cloud Computing, 3(2):36–45, 2016.
- [51] Ismini Psychoula, Deepika Singh, Liming Chen, Feng Chen, Andreas Holzinger, and Huan-sheng Ning. Users’ privacy concerns in iot based applications. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), pages 1887–1894. IEEE, 2018.
- [52] Ramprasad Ravichandran, Michael Benisch, Patrick Gage Kelley, and Norman M Sadeh. Capturing social networking privacy preferences. In International symposium on privacy enhancing technologies symposium, pages 1–18. Springer, 2009.
- [53] Norman Sadeh, Jason Hong, Lorrie Cranor, Ian Fette, Patrick Kelley, Madhu Prabaker, and Jinghai Rao. Understanding and capturing people’s privacy policies in a mobile social networking application. Personal and Ubiquitous Computing, 13(6):401–412, 2009.
- [54] Mohamed Seliem, Khalid Elgazzar, and Kasem Khalil. Towards privacy preserving iot environments: a survey. Wireless Communications and Mobile Computing, 2018, 2018.
- [55] Parvaneh Shayegh and Sepideh Ghanavati. Toward an approach to privacy notices in iot. In 2017 IEEE 25th International Requirements Engineering Conference Workshops (REW), pages 104–110. IEEE, 2017.
- [56] M Shobana and S Rathi. Iot malware: An analysis of iot device hijacking. 2018.
- [57] Tanuja Singh and Mark E Hill. Consumer privacy and the internet in europe: a view from germany. Journal of consumer marketing, 2003.

- [58] Daniel J Solove. Introduction: Privacy self-management and the consent dilemma. Harv. L. Rev., 126:1880, 2012.
- [59] Nili Steinfeld. “i agree to the terms and conditions”:(how) do users read privacy policies online? an eye-tracking experiment. Computers in human behavior, 55:992–1000, 2016.
- [60] Xavier Tannier. Webannotator, an annotation tool for web pages. In LREC, pages 316–319, 2012.
- [61] Sabine Trepte, Leonard Reinecke, Nicole B Ellison, Oliver Quiring, Mike Z Yao, and Marc Ziegele. A cross-cultural perspective on the privacy calculus. Social Media+ Society, 3(1):2056305116688035, 2017.
- [62] Zhi-Kai Zhang, Michael Cheng Yi Cho, Chia-Wei Wang, Chia-Wei Hsu, Chong-Kuan Chen, and Shiuhpyng Shieh. Iot security: ongoing challenges and research opportunities. In 2014 IEEE 7th international conference on service-oriented computing and applications, pages 230–234. IEEE, 2014.
- [63] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home iot privacy. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–20, 2018.