



# Unveiling Users' Security and Privacy Concerns Regarding Smart Home IoT Products from Online Reviews

TAUFIQ ISLAM PROTICK, North Carolina State University, Raleigh, United States

AAFAQ SABIR, North Carolina State University, Raleigh, United States

SB ABHINAYA, North Carolina State University, Raleigh, United States

AIDEN BARTLETT, North Carolina State University, Raleigh, USA

ANUPAM DAS, North Carolina State University, Raleigh, United States

The Internet of Things (IoT) has revolutionized the global market with lifestyle products such as fitness trackers (FT), smart home speakers (SHS), and surveillance and security camera systems (SSCS). While offering convenience, these products also introduce potential security and privacy (S&P) risks to buyers, often going unnoticed. Consumers' incomplete mental models of the risks involved and the information asymmetry between buyers and sellers only add to the problem. Understanding consumer concerns in online product reviews can play a crucial role in bridging the gap of such information asymmetry. By establishing a balanced flow of information between buyers and sellers, manufacturers can leverage genuine concerns expressed in reviews to enhance product features while educating users about misinformation in reviews. In this study, we collected FT, SHS, and SSCS product reviews from three Amazon domains: the US, the UK, and India. Using a keyword-based search method focused on S&P concerns, we discovered a considerable number of reviews expressing notable concerns regarding security and privacy. Our qualitative analysis revealed that *data security* is a common concern across all product types. Further, our quantitative analysis exposed significant geographic variations, with the concern ratio being higher in the US than in the UK for all device types and higher than in the Indian domain for security cameras. These findings highlight the need for tailored security measures and user awareness campaigns in different parts of the world to address the identified concerns effectively.

CCS Concepts: • **Security and privacy** → **Human and societal aspects of security and privacy**.

Additional Key Words and Phrases: IoT devices, Mixed-method analysis, Security and Privacy concerns

## 1 Introduction

The Internet of Things (IoT) market is expanding rapidly, with 29 billion connected devices anticipated by 2027 [10]. Among the IoT devices, fitness trackers, smart home speakers, and surveillance and security camera systems have gained significant popularity [25, 35, 36]. The capability for monitoring health with ease using wearable fitness trackers [4, 35], automating tasks using voice-controlled commands through smart speakers [16, 36], and enhanced security for household and office environments through smart surveillance cameras [25] have led to rapid growth in the consumer adoption of these IoT devices.

---

Authors' Contact Information: Taufiq Islam Protick, North Carolina State University, Raleigh, North Carolina, United States; e-mail: tprotic@ncsu.edu; Aafaq Sabir, North Carolina State University, Raleigh, North Carolina, United States; e-mail: asabir2@ncsu.edu; SB Abhinaya, North Carolina State University, Raleigh, North Carolina, United States; e-mail: asrivid@ncsu.edu; Aiden Bartlett, North Carolina State University, Raleigh, North Carolina, USA; e-mail: aebartl3@ncsu.edu; Anupam Das, North Carolina State University, Raleigh, North Carolina, United States; e-mail: anupam.das@ncsu.edu.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, or post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2834-5533/2024/8-ART

<https://doi.org/10.1145/3685929>

Although these products offer convenience, they also come with several security and privacy (S&P) issues. Apps used for fitness tracking have been found to reveal location data [28]. The presence of continuously listening microphones in smart home speakers such as Amazon Echo and Google Home have raised privacy concerns [48]. Reported incidents of unauthorized recording and sharing of private conversations by an Amazon Alexa device further highlight privacy issues with IoT devices [43]. Some smart surveillance camera systems have been found to have weak security implementations [25]. Instances of hacking into vulnerable devices, such as the Foscam baby monitor [65] or the VTech baby monitor [45], serve as concrete evidence underscoring the gravity of S&P risks associated with these systems [63].

S&P issues in IoT devices undermine trust and create skepticism among consumers. Concerns about online fraud, misuse of personal information, and the perceived lack of control over personal data contribute to this skepticism [57, 66]. Consumers become more concerned about S&P after exposure to media reports, opinions from friends, or unexpected device behavior [32]. A survey conducted in the UK reported security information as the third most important type of information sought—after cost and functionality—when making a purchase decision [6]. However, finding S&P-related information before purchasing IoT products remains a challenge [32]. To address this issue, researchers have proposed the development of a comprehensive S&P labels for IoT products, incorporating expert input and end-user opinions [29]. Despite these efforts, the practical implementation of influential S&P labels is yet to be widely adopted. Outside the field of IoT, (in iOS and Android apps), findings indicate that even the adoption of such labels does not adequately answer people’s inquiries related to privacy [75].

Users’ lack of awareness about the exact set of S&P issues in a particular IoT device and the consequent lack of their adoption of required protective measures make users susceptible to several risks. Users often neglect modifying default privacy settings and disregard the privacy policies of IoT products [37, 77]. User awareness can be enhanced by leveraging existing consumer product reviews, which often provide valuable insights regarding existing S&P issues for an IoT device. Given that prospective buyers look through reviews when making a purchase decision [76], extracting S&P concerns from reviews can aid decision-making, benefiting both the consumers and the market. While reviews usually reveal genuine problems, they may contain misinformation as well. The dissemination of consumers’ S&P perceptions in a public forum creates a channel between the user and the manufacturer, akin to the concept of a user-to-developer channel advocated in the domain of app review analysis [42]. Eliciting user concerns from product reviews enables manufacturers to address those concerns, clarify misconceptions, and build trust, thereby enhancing IoT adoption.

Prior works focusing on understanding users’ S&P perceptions are specific to a type of IoT device or a particular geographic location [16, 18, 27]. Other studies have employed surveys and interviews to understand consumer perceptions about a broad range of IoT devices [31, 40, 71]. Li et al. leveraged threads on Reddit to understand S&P attitudes of smart home users [49]. Vetrivel et al. investigate how consumer reviews of smart routers and hubs reflect well-known S&P concerns [69]. They examined over 80K reviews from Amazon for four device types known to have been infected by the Mirai botnet. All these techniques, while effective, are targeted toward a specific device category. In contrast, we focus on three distinct categories of IoT devices with a high potential for being invasive: fitness trackers (FT), smart speakers (SHS), and security cameras (SSCS). Fitness trackers collect highly sensitive health-related information such as user weight, blood pressure, and steps taken, which they can use to detect the health of users’ heart and lungs, menstrual cycle, and sleep patterns [11]. The raw audio data that smart home speakers collect can be used to infer the age, gender, accent, and emotional state of a speaker using speech signal processing [44, 74]. Smart surveillance and security cameras collect visual data that can reveal information about user’s physical location, putting them at risk of physical harm [27]. By delving deeper into the corpus of user reviews, we identify S&P perceptions specific to these three classes of IoT devices.

In this paper, we seek to answer the following research questions. **RQ1: How can we leverage consumers’ IoT product reviews to identify S&P concerns?** We compile an *enhanced-keyword-set* and use it in a search-based

pipeline to extract reviews from Amazon, a substantial proportion of which express S&P concerns. **RQ2: *What S&P concerns emerge from consumers' IoT product reviews?*** We conduct thematic analysis [23] with reflexive coding approach [24] on reviews expressing S&P concerns to identify key themes across three device types. **RQ3: *How do IoT consumers' S&P concerns vary across geographic locations?*** We perform statistical tests to quantify differences in S&P concern ratios among the US, the UK, and India, providing insights into user concerns across diverse geo-cultural and regulatory contexts.

In summary, we make the following contributions:

- To the best of our knowledge, our research marks the initial exploration of security and privacy (S&P) concerns related to three lifestyle-oriented IoT device categories—fitness trackers (FT), smart speakers (SHS), and security cameras (SSCS)—using consumer product reviews. We selected these devices due to the tangible risks associated with their everyday usage [28, 43, 45, 48, 65]. We use an enhanced set of keywords, referred to in this paper as the *enhanced-keyword-set*, that was derived using *inductive coding*, enabling us to extract meaningful Amazon reviews expressing S&P concerns for each IoT device type (30% in FT, 43% in SHS, and 27% in SSCS).
- We identify 14 S&P-related themes stemming from concerns regarding data security, technical issues with the device or the accompanying apps, affiliations of the device manufacturer with various entities, and superficial or unfounded opinions or misconceptions.
- We analyze the distribution of S&P concerns for these devices across the US, the UK, and India domains. We observe that reviews from the US consistently exhibited higher concern ratios than reviews from the UK across all device types. Additionally, concerns for security cameras were significantly higher in the reviews from the US domain than those from the Indian domain.

## 2 Related work

**Consumers' data awareness, mental model, and trust on IoT products.** Researchers have identified that while fitness tracker users are confident about data collection, they are uncertain about its usage and are more willing to share data with friends than strangers [37]. Aktypi et al. examined information sharing on social networks and the potential leakage of personal data from fitness-tracking devices, revealing that participants paid little attention to the associated risks [18]. Lidynia et al. found that long-term storage location and possible recipients of the collected data are perceived as privacy sensitive factors for users [50]. Velykoivanenko et al. concluded that participants were concerned about inference of their personality trait, socioeconomic status, political views, alcohol or tobacco consumption, or sexual activity from their fitness tracker devices [68]. Becker et al. found that participants' privacy concern depended on the recipients of PHI (Protected Health Information) as they were reluctant if the PHI is provided to third-party but were relatively open-minded if the manufacturers used this information to improve their products [21]. Vitak et al. found that majority of their respondents (73%) did not know whether Fitbit/Jawbone sold their data, and 66% were not sure who owned their data [70]. Schomakers et al. studied the privacy concerns for e-Health technologies: fitness trackers and remote monitoring of implanted cardiac devices and found that privacy concerns for fitness trackers are significantly higher than other remote monitoring devices [61]. These studies highlight the importance of informing users about S&P risks in FT products and motivated us to include fitness trackers to the device categories that we focus on in this work.

Malkin et al. surveyed users of smart speakers (Amazon Alexa and Google Home) and found that many users were unaware that their recordings were permanently stored, with only a small fraction of users reviewing or deleting their recorded interactions [52]. Security cameras pose the risk of data collection by intruders exploiting device vulnerabilities, often unknown to the consumer. For instance, exploiting a security vulnerability in a Foscam baby monitor enabled remote internet monitoring from anywhere in the world [65]. This security

breach occurred despite researchers disclosing the vulnerability to the manufacturer a few months prior to the incident [63].

Prior research on the privacy practices of users of activity-tracking devices reveals that privacy is a flexible concept resulting from users' negotiation between the benefits and harms of disclosing personal information [58]. The results of an interview study of 17 smart speaker users and 17 non-users showed that non-users lacked trust in smart speaker companies, while the users had few concerns [48]. This finding hints at the incorrect mental model of the smart speaker users with an incomplete understanding of the privacy implications of the product. Such results were also supported by Abdi et al. in their study, which considered four use-case scenarios for smart speakers with distinctive architectural elements and stakeholders [16]. Manilkonda et al. conducted a survey and found that while users were concerned about privacy, they were still optimistic about using smart speakers [53]. These findings reveal that users have varying levels of concern about smart speakers. Zeng et al. found smart cameras as one of the most popular devices in the context of a smart home [72]. They also found that "physical security" was the most common asset the smart camera owners wanted to protect; however, participants worried that smart cameras could also put their physical security at risk if compromised. Another study found audio and video access via smart cameras to be the most frequent concern among their survey participants [41]. Others studied the privacy perception of smart cameras and showed that the context and use case could significantly affect privacy perceptions. This suggests that even users from similar demographics might have very different concerns [27, 73].

**Eliciting privacy concerns from product reviews.** To better inform customers about S&P aspects of IoT devices at the time of purchase, various "signals" have been proposed and tested by prior research [60, 64]. These signals can be broadly divided into two types: manufacturer controlled (e.g., security labels) and non-manufacturer controlled (e.g., customer reviews) [17]. Much research has been done on manufacturer-controlled signals, specifically security labels, to measure their effectiveness in informing customers about S&P concerns with IoT devices [29, 30, 33, 54]. These studies concluded that security labels are informative but not yet widely used. Moreover, as they are self-certified labels, they do not reflect the real-world consumer experiences regarding S&P or even provide the actual security performance of the devices [38]. Therefore, considering the limitations of security labels and research suggesting buyers trust customer reviews at the time of purchase [76], researchers have also focused on consumer-controlled signals, i.e., product reviews [69].

Prior research has investigated product reviews to extract customer concerns regarding various aspects of technology. Linden et al. conducted a mixed-method, comparative study to examine if people had a different perception of privacy for pet wearables compared to human wearables [67]. In their study, they extracted privacy concerns found in around 8K product reviews of pet wearables, contrasting them with 20K reviews of human wearables. The findings suggested that very few concerns about pet or human wearables are voiced. It was found that only 0.06% (5 of 8,038) and 1.4% (29 of 20,431) of reviews on pet and human wearables, respectively, mentioned privacy concerns.

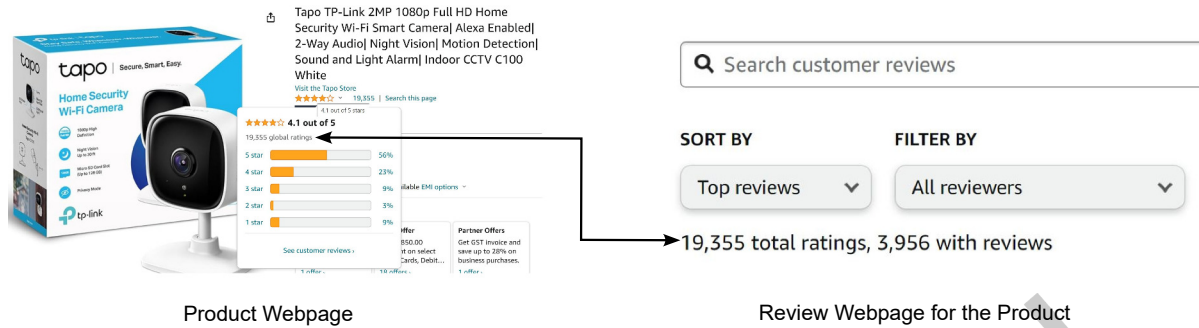
One of the first efforts to understand how consumers discuss S&P concerns regarding smart home assistant devices within product reviews was investigated by Fruchter et al. [36]. They created a dictionary with 53 keywords (specified in Table 6) to identify S&P-related reviews in their work. They found 109,536 reviews using their selected keywords. Two coders read the tagged reviews to find the true positives, i.e., the ones with S&P concerns. Their finding indicates that 2% (2,237 of 109,536) of the total reviews contain S&P concerns. They found three major themes of concern: '*data collection and scope*', '*creepy device behavior*', and '*violations of personal privacy thresholds*'. The 53 keywords they chose comprised a broad set of S&P-related terms sourced from prominent S&P-related incidents and media coverage of such incidents. However, their keywords are too broad to identify S&P concerns related to specific device categories, as we will demonstrate later on. Vetrivel et al. examined 83,686 reviews from Amazon sites representing six geographic regions for four device types vulnerable to Mirai botnets [69]. With the help of topic modeling and manual coding, they concluded that one in ten reviews

**Table 1.** Tabular representation of the aspects in prior works vs. our work.

	Fruchter et al.[36]	Linden et al. [67]	Vetrivel et al. [69]	Our work
Device categories chosen	Smart speakers	Pet and human wearables	Surveillance systems, routers, hubs, set-top boxes, once vulnerable routers	Fitness trackers, Smart home speakers, Security cameras
Nature of the devices	Endpoints	Endpoints	Mostly gateways, one endpoint	Endpoints
Choice of keywords	Common terms were discussed and identified, and keywords from significant privacy and security events were integrated into the final dictionary.	Keywords were derived from Motti and Caine's privacy analysis [55] in wearables, using their coded review fragments followed by keyness analysis.	The keywords used for searching were based on prior work [39], aiming to reflect how users express concerns rather than technical terminology often used in research.	Compiled from empirical evidence, i.e., reading actual product reviews and picking up words or phrases from parts of the reviews where concerns were mentioned.
Key-phrase used	Yes	Yes	No	Yes
Cross-geographic analysis	No	Yes	Yes	Yes
Outcome	Concerns related to S&P account for 2% of the total reviews. Data collection and scope, creepy device behavior, and violations of personal privacy thresholds were the major concerns found.	Emotional drivers and functionality needs overshadow privacy concerns in pet wearables. Addressing this requires understanding consumer behavior and updating data protection laws.	One in 10 reviews mentions security and privacy issues, varying by country. Seven themes, including technical language and friction with security steps had been identified.	Thematic analysis shows data security is a common concern across all product types. Humor and sarcasm are used to express concerns. Geographically, concerns are higher in the US than in the UK and India.

(9.8%) mentioned S&P issues. However, three of the four device types chosen in their study: routers, hubs, and set-top boxes, are considered more as IoT gateways (connecting devices to the Internet) than endpoints (with which a user directly interacts).

**Distinction from prior work.** Our work stands out due to a nuanced blend of four critical factors. Firstly, our choice of device categories stems from empirical evidence within the review corpus, focusing on the most popular and widely reviewed products while avoiding known vulnerabilities, e.g., devices vulnerable to Mirai Botnet attacks [69]. This allows us to report on people's S&P perception about devices that do not have widely known vulnerabilities. Secondly, our specialized keyword selection, expertly curated from real reviews, sidesteps the noise and false positives generated by generic keywords, allowing us to get to the niche where we find moderately large but quality data rife with concerns. We made a contrast between the efficacy of our *enhanced-keyword-set* and the keywords used by Fruchter et al. [36] in finding S&P concerns when they were used to search for product reviews for the same set of products. Our keyword set extracted more than double the proportion of concerns for smart home speakers (47.7% vs. 17.16%) in the retrieved reviews that matched the keywords. Thirdly, our emphasis on human interaction with devices distinguishes us from studies focusing mostly on IoT gateways [69], ensuring an understanding of people's concerns about devices they directly interact with. Lastly, although some device categories were previously explored individually (smart home assistants in [36], and pet and human wearables in [51]), our work combines both (FT and SHS), along with SSCS to explore and compare the user-oriented privacy concerns in lifestyle IoT devices. Overall, qualitative analysis in all levels of our methods gives us high-quality



**Fig. 1.** An example of how we connect one piece of information with the other, i.e., the global ratings (left) with the ratings with reviews (right).

labeled data that would allow us to utilize them to train a machine learning model for future automation. In terms of outcomes, we find humor or sarcasm in people's expressed concerns that were not explicitly mentioned by these works [36, 67, 69]. However, the use of sarcasm to convey these concerns finds support in another line of research [22] that studied the extent to which participants talked about security and privacy on social media platforms. In terms of similarities in findings, our study corroborates to the findings from Vetrivel et al. [69] since they also found the percentage of S&P reviews to be the highest in the US domain. Table 1 summarizes these distinctions.

### 3 Methodology

#### 3.1 Reasons for choosing FT, SHS, and SSCS as product categories

These three categories—fitness trackers, smart home speakers, and security cameras—were selected due to specific privacy concerns related to their functionalities. Fitness tracking apps, for example, have been shown to inadvertently disclose users' locations [28]. Additionally, smart home speakers like Amazon Echo and Google Home, which feature constantly active microphones, have raised significant privacy apprehensions [48]. Moreover, incidents involving unauthorized recording and sharing of private conversations by Amazon Alexa devices have heightened concerns about IoT device privacy [43]. Furthermore, certain smart surveillance camera systems exhibit vulnerabilities in their security implementations [25]. Notably, instances of hacking into devices like the Foscam baby monitor and the VTech baby monitor serve as concrete examples illustrating the substantial security and privacy risks associated with these systems [45, 63, 65]. The three types of IoT devices that we selected are actively used by users as everyday lifestyle products (FT for monitoring physical activity, SSCS to monitor important places, and SHS as virtual assistants) unlike hubs or routers considered in some of the prior works [69]. Therefore, users also observe or speculate privacy concerns based on their personalized usage experience thus motivating them to write a review on Amazon which is our vantage point of eliciting users privacy concerns.

#### 3.2 Data collection, processing, and analysis

Figure 3 outlines our comprehensive data collection and analysis procedure with the following steps: gathering data from Amazon stores (Steps 1, 2); subsampling the collected data, and employing an inductive coding approach to construct an *enhanced-keyword-set* (Steps 3-5); using the *enhanced-keyword-set* in web stores to find reviews containing them (Steps 6, 7); subsequent human coding to label reviews with S&P concerns and filter concerning reviews (Steps 8, 9); thematic coding of the reviews with S&P concerns (Steps 10, 11); finally, conducting statistical tests for cross-geographic analyses on labeled reviews (Steps 12, 13). Further details are given below.

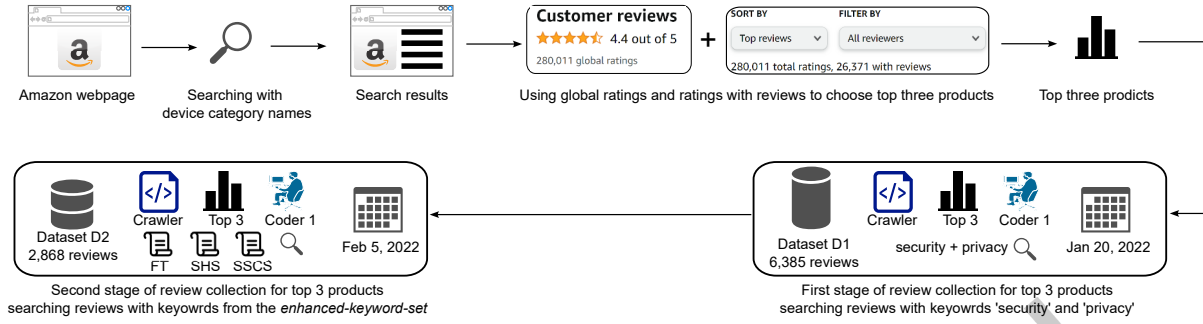


Fig. 2. An illustration of the product choosing criterion and the timestamps for running the crawler for data collection.

**Platform and web domains.** We chose Amazon as our target platform to collect IoT product reviews as it is one of the largest e-commerce platforms available across different countries [62]. In this platform, we chose three country-specific web domains from which we collected product reviews: `amazon.com`, `amazon.co.uk`, and `amazon.in`. These three domains were chosen because not only do they contain reviews in English but also cover three distinct geographic regions. The chosen device types for analysis were fitness trackers (FT), smart home speakers (SHS), and surveillance and security camera systems (SSCS).

**Review collection.** To collect reviews, we searched for each device type on the homepage of each Amazon domain and selected the top three products based on global ratings. Global ratings is the total number of Amazon ratings with or without reviews [12]. From empirical evidence, we saw that the higher the global ratings, the more the reviews. Also, global ratings have an impact on the average product rank [15]. Therefore, for choosing products, it was used as a metric that increases our likelihood of obtaining more reviews. Figure 1 shows how we connect the two pieces of information, i.e., the global rating and the number of ratings with reviews for a security camera. Figure 2 shows that this criterion was used to choose the top three products for each category and domain. Finally, on January 20, 2022, we collected the first dataset  $\mathcal{D}_1$  searching with the keywords 'security' and 'privacy'. Later, for the same set of products we collected the dataset  $\mathcal{D}_2$  searching with the keywords from the *enhanced keyword set*.

**Product selection.** Although we aimed for consistent product selection across domains, variations occurred when certain products were unavailable or less popular in a domain. Products in the same category had similar features (FT: Time Display, Sleep Monitor, Activity Tracker, GPS, Heart Rate Monitor; SHS: Voice Control with Alexa, Speaker, Smart Home Control; SSCS: Live View, Motion Detection, Recording Footage, Storage in the cloud), with a slight difference between Echo Dots and the Echo Show, where the Echo Show, in addition, has a display screen and a built-in camera. Table 2 shows the selected products in each domain. In this paper, the flags of the US, UK, and India represent `amazon.com`, `amazon.co.uk`, and `amazon.in`, respectively.

It is to note that due to the involvement of human analysis at each step of our study, we could not choose a very large dataset that would prohibit human analysis due to human fatigue and errors associated with analysis. This is why we restricted ourselves to a moderately large dataset. We do this by choosing the top three products based on global ratings, which would increase our chances of getting a moderately large number of reviews since more global ratings imply more reviews with ratings, too. An important side effect (which was not intentional) of such a choosing criterion is getting products in the FT and SHS categories from the same vendor (Fitbit and Amazon Echo, respectively). This may limit the generability and applicability of our findings for some other brands.

**Curating datasets.** We automated our data collection process using Selenium for web crawling and BeautifulSoup for parsing HTML files. Data collection was conducted in two phases. Dataset  $\mathcal{D}_1$  is the dataset of product reviews



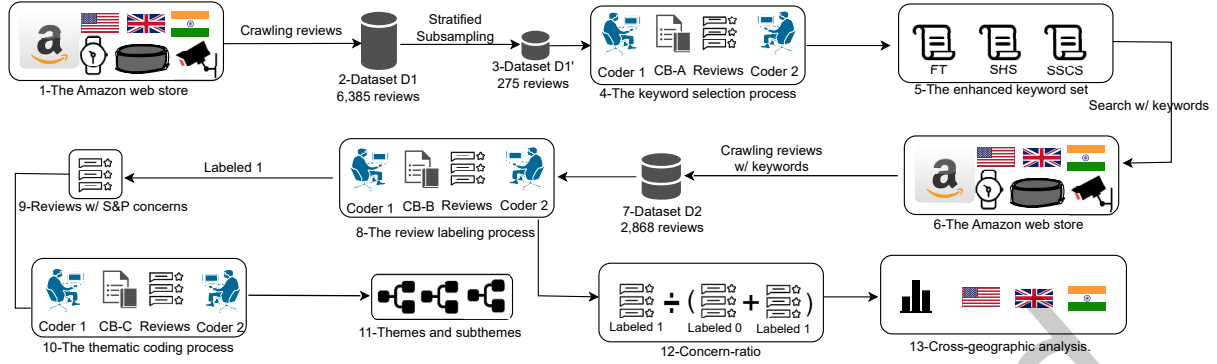


Fig. 3. A high-level overview of our analysis of IoT product reviews. Here, CB refers to Codebook.

Table 2. The list of products whose reviews are collected across different geographic regions. Different ASINs are used to identify different products.










Device Type	FT			SHS			SSCS						
Product	Fitbit Versa 2	Fitbit Charge 4	Fitbit Inspire 2	Echo Dot (3rd Gen)	Echo Dot (4th Gen)	Echo Show 8	Blink Mini	Blink Outdoor	Wyze Cam	TP Link Tapo	TP Link 2Mp	Mi 360	realme 360
ASIN	B07TYNMRZG B07TVC2KLW	B084CQ41M2	B08DFGPTSK	B07FZ8S74R B07PJV3JPR B07PFMF9P9	B07XJ8C8F5 B084DWXCZXZ B084KSRFXJ	B07PF1Y28C B07SMN7FYZ B07SMNPPCGK	B07X6C9RMF B07X37DT9M B086DKSHQ4	B088CZW8XC	B08E59YH7W	B07XR2NWXZC	B083V41T6M	B07HJD1KH4	B08LZH4FYQ
Domain													

collected in the first phase of our study. For each product in Table 2, we did a keyword-based search on the Amazon web store for reviews with our first set of keywords  $\{privacy, security\}$  and collected  $\mathcal{D}_1$ . This means each review in  $\mathcal{D}_1$  is a review that contained the keyword  $k \in \{privacy, security\}$  in the *title* or the *content* of the review. Dataset  $\mathcal{D}_2$  is the dataset of product reviews collected in the second phase of our study. The reviews were collected from the same product list as of  $\mathcal{D}_1$ . In this phase, we did a keyword-based search for reviews with our second list of keywords, the *enhanced-keyword-set*. We describe our methodology for compiling the *enhanced-keyword-set* in §3.5. Unlike the set of generic search words we used in phase one, i.e.,  $\{privacy, security\}$ , the *enhanced-keyword-set* is device-type specific. Table 3 contains the details of datasets  $\mathcal{D}_1$  and  $\mathcal{D}_2$ . For each review, we collected ten types of **metadata**: *title*, *content*, *date*, *variant* (e.g., the color of a product), *images* (URL of an image file if the review had any attached image), *verified* (if the review came from a verified purchaser), *author* (with user's geographic location), *rating*, *product*, and *URL* (the link to the page where we found the review when we made our search).

Data processing involved removing reviews with empty bodies, eliminating duplicates, and filtering out non-English reviews. For  $\mathcal{D}_1$ , we collected 6,941 reviews (FT: 218, SHS: 2,148, SSCS: 4,575). After filtering, we obtained 6,385 reviews (FT: 78, SHS: 2,028, SSCS: 4,279), resulting in an 8.01% reduction (FT: 64.22%, SHS: 5.59%, SSCS: 6.47%). Similarly,  $\mathcal{D}_2$  had 5,184 reviews (FT: 534, SHS: 3,168, SSCS: 1,482), reduced to 2,868 reviews (FT:



**Table 3.** Dataset  $\mathcal{D}_1$  (left) of reviews containing keyword  $k \in \{\textit{privacy}, \textit{security}\}$ . Dataset  $\mathcal{D}_2$  (right) of reviews containing keywords belonging to the *enhanced keyword set*.

Device Type	$\mathcal{D}_1$ : Collected on Jan 20, 2022									$\mathcal{D}_2$ : Collected on Feb 5, 2022			
	Total # of reviews	Keyword: “privacy”	Collected from			Keyword: “security”	Collected from			Total # of reviews	Collected from		
													
FT	78	14	9	5	–	64	56	5	3	142	100	33	9
SHS	2028	545	377	152	16	1483	1060	410	13	1845	1109	689	47
SSCS	4279	118	56	41	21	4161	2974	692	495	881	582	164	135
Total	6385									2868			

142, SHS: 1,845, SSCS: 881), a 48.68% reduction (FT: 73.4%, SHS: 41.76%, SSCS: 40.55%). This reduction primarily resulted when the same review was found while searching with different keywords.

In this paper, we represent a review (or a part of it) with its device type followed by its serial number in the processed dataset  $\mathcal{D}_2$ . For example, FT-206 would mean the 206<sup>th</sup> review for FT category in  $\mathcal{D}_2$ . If any reference is made from other datasets, the name of the dataset will be suffixed in the end, e.g., FT-100- $\mathcal{D}_1$ . The reviews are represented verbatim with no typos corrected. If a review is too long, the relevant part is represented with a leading and/or trailing ellipsis, i.e., "...".

### 3.3 Definition of an S&P concern and Creation of codebooks

In our study, we define an S&P concern as follows: *An S&P concern is a statement encompassing a range of perspectives, including comments, opinions, skepticism, and real-life incidents. It explores critical considerations related to confidentiality, integrity, authentication, secrecy, anonymity, solitude, and other relevant factors that impact a consumer's security and privacy perspectives.* The design of our mixed-method analysis is centered around searching for product reviews that match any of the aspects in this definition. The data analysis at different stages generated three codebooks as an outcome. Table 4 provides the instructions for generating each codebook (more details available in Appendix A, B, and C). The codes of all three codebooks can be anonymously viewed.<sup>1</sup> Codebook A gave us the *enhanced-keyword-set* to do an advanced search for reviews with those keywords. Codebook B produced labels for a review to either have concerns or no concerns. Codebook C gave us fine-grained subthemes to categorize reviews with concerns for the thematic analysis. When two coders collaborated, the codebooks were revised with insights contributed by each coder after an independent coding iteration.

### 3.4 Subsampling $\mathcal{D}_1$ to construct $\mathcal{D}'_1$




We read a stratified sample of  $\mathcal{D}_1$ , called  $\mathcal{D}'_1$ , to create the *enhanced-keyword-set*. The number of reviews containing the keywords 'security' vs. 'privacy' varied across device types: FT (64 vs. 14), SHS (1,483 vs. 545), and SSCS (4,161 vs. 118). To handle the disproportion, our goal was to include 300 reviews in  $\mathcal{D}'_1$ , with stratification performed by device type (100 each), web domains (split as 34-33-33), and keywords (16/17 each for 'security' and 'privacy'). Any uneven splits were resolved by adding an extra review from the dataset with the largest proportion. All available data were included if the available data for a domain, device type, or keyword was smaller than the expected subsample size. Simple random sampling [14] was employed during stratification since it ensures every individual review in  $\mathcal{D}_1$  to have an equal chance of being selected, promoting unbiased and representative results. Table 5 provides more details on  $\mathcal{D}'_1$ .

<sup>1</sup>The codebooks are available here [https://osf.io/x4npj/?view\\_only=dd37575bf1d9414c9e463fd1dd5cadaf](https://osf.io/x4npj/?view_only=dd37575bf1d9414c9e463fd1dd5cadaf)

**Table 4.** Rules followed in each task of the coding process.

Task	Rules followed
A. Keyword Selection	<ul style="list-style-type: none"> <li>• Coders chose keywords from reviews that addressed security and privacy concerns.</li> <li>• Keywords were selected only from the sections of a review where security and privacy concerns were discussed.</li> <li>• Keywords were included in their original form without any modifications. For example, “invades” would be kept as is rather than being reduced to “invade.”</li> <li>• Keywords were selected even if they consisted of multiple words concatenated together, such as “card information,” “beware of privacy,” and “phone number.”</li> </ul>
B. Review Labeling	<ul style="list-style-type: none"> <li>• Reviews labeled as ‘1’ (S&amp;P concern) meet one or more of the following conditions: <ul style="list-style-type: none"> <li>– The review discusses the negative security and privacy aspects of the product.</li> <li>– The review sarcastically addresses security and privacy concerns while acknowledging their existence.</li> <li>– The review acknowledges privacy violations but still expresses an intention to use the product.</li> </ul> </li> <li>• Reviews labeled as ‘0’ (no S&amp;P concerns) fall into one of the following categories: <ul style="list-style-type: none"> <li>– The review is ambiguous and does not explicitly mention any security or privacy concerns.</li> <li>– The review does not discuss any security or privacy aspects of the product.</li> <li>– The review comments on the security and privacy aspects in a positive or neutral manner.</li> </ul> </li> </ul>
C. Thematic Analysis	<ul style="list-style-type: none"> <li>• The coding process was inductive, starting with an empty set of codes.</li> <li>• Codes were developed in a bottom-up manner while analyzing the reviews.</li> <li>• Each concern review was carefully examined, including both the title and the body.</li> <li>• Relevant parts of the concerns were identified and highlighted as <i>excerpts</i>.</li> <li>• <i>Excerpts</i> were initially assigned to specific low-level subthemes.</li> <li>• An <i>excerpt</i> could be assigned to multiple subthemes if it was relevant to each of them.</li> <li>• After reading a dataset, similar subthemes were grouped together and categorized under high-level themes.</li> <li>• Self-reflection was practiced to reduce bias, interpret nuances, and update own mental model before moving on to do thematic coding on the next dataset.</li> </ul>

**Table 5.** The initial number (left) of reviews with keyword ‘privacy’ vs. ‘security’ in  $\mathcal{D}_1$ . Dataset  $\mathcal{D}'_1$  (right) subsampled from  $\mathcal{D}_1$  to balance the disproportion and read to derive the *enhanced-keyword-set*.

Device Type	Initial quantity and proportion of reviews in $\mathcal{D}_1$ with keyword			Stratified Subsampling in $\mathcal{D}'_1$		
	Privacy	Security	Sample size	 (Privacy + Security)	 (Privacy + Security)	 (Privacy + Security)
FT	14 (17.95 %)	64 (82.05%)	78	10 (5+5)	65 (9+56)	3 (0+3)
SHS	545 (26.87%)	1483 (73.13%)	97	34 (17+17)	34 (17+17)	29 (16+13)
SSCS	118 (2.76%)	4161 (97.24%)	100	34 (17+17)	34 (17+17)	32 (16+16)

### 3.5 Compiling the enhanced-keyword-set by manual review of $\mathcal{D}_1'$

Two researchers with experience in IoT security and privacy independently read the title and content of each review in  $\mathcal{D}_1'$  to compile the *enhanced-keyword-set*. The initial agreement percentages for keyword selection across the three device types were: FT: 65.30%, SHS: 78.95%, SSCS: 80%. The relatively lower agreement observed in FT can be attributed to two primary reasons. Firstly, it stems from the inherent challenge of identifying suitable candidate words for FT as in FT-211- $\mathcal{D}_1$  below.

*FT-211- $\mathcal{D}_1$  (searched with 'security'): "The wrist strap does not **securely** fasten to the watch, and I lost my watch after two days."*

Many reviews like FT-211- $\mathcal{D}_1$  had been found in our first round of data collection where the keyword "security" was used to search for reviews. Most reviews retrieved by the search process were not relevant as they contain complaints about the wrist strap not "securely" fitting to the wrist or the watch. Secondly, independent coding for each category was interspersed with conflict resolution. Beginning with FT keywords, resolving conflicts at this stage led to fewer conflicts while coding SHS and SSCS reviews. Words present in both coders' sets were directly included in the *enhanced-keyword-set*. In cases of disagreement, each coder presented their reasoning for the disputed words. If the justification was mutually accepted, the word was added to the *enhanced-keyword-set*; otherwise, it was excluded. Figure 5 shows an illustrated example of this step. During our compilation and search with the *enhanced-keyword-set*, we deliberately avoided stemming and lemmatization for two reasons. Firstly, Amazon's search portal does not always retrieve reviews containing the exact keyword. Therefore, we assume there is some implicit regex check that runs on their end. For example, the review FT-211- $\mathcal{D}_1$  was found when searched with the keyword 'security'. However, the review does not contain any words that exactly match our search word. Rather it contains the word 'securely'. Secondly, using stemmed or lemmatized forms of a keyword most often turns a word into its generic form, which may retrieve a lot of noisy reviews: the ones we hypothesize are not conducive to what we intend to find in our study. On the search portal for Amazon, when we used the stemmed form of the word security, i.e., 'secur' we obtained reviews containing 'secure' but it was related to the device's strap.

*(Searched with 'secur', the stemmed form of 'security' for Fitbit Versa 2, ASIN: B07TYNMRZG): "Overall it does exactly what is required and easily. Only complaint is that the strap keeper is a little loose and keeps slipping around. I have added another tighter keeper to **secure** the strap in place better. For the price it is still great value."*

### 3.6 Collecting data using the enhanced-keyword-set

Next, we utilized our *enhanced-keyword-set* to gather reviews representing  $\mathcal{D}_2$ . As mentioned in §3.5 Amazon's search portal inherently conducts specific regex checks that assist in extracting reviews containing variations of the keywords. Also, we deliberately avoided stemming and lemmatization during the search because they tend to overly generalize search words, inadvertently including reviews with unintended keywords and introducing noise. This is why we opted for the original keywords without their root forms, as stipulated as an instruction listed in Table 4 (the third bullet point in A. Keyword Selection). Since each review in  $\mathcal{D}_2$  should now contain any keywords from the *enhanced-keyword-set*,  $\mathcal{D}_2$  is a dataset representative of the reviews mentioning any words from the *enhanced-keyword-set*.

### 3.7 Labeling reviews in $\mathcal{D}_2$

Reviews within  $\mathcal{D}_2$  were labeled as either '1' (indicating the presence of S&P concerns) or '0' (indicating no S&P concerns), adhering to the guidelines outlined in the instructions to generate Codebook B. The coders independently assigned labels based on the review's title and content, following the rules detailed in the row "B. Review labeling" in Table 4. The proportion of verified reviews in each category is as follows: FT: 114/142 (80.28%), SHS: 1746/1845 (94.63%), SSCS: 684/881 (77.64%). Since our primary objective was to assess the efficacy of the keywords in identifying S&P concerns, we categorized all the reviews in  $\mathcal{D}_2$  without excluding the unverified

ones, acknowledging that both verified and unverified reviews are representatives of the reviews containing any words from the *enhanced-keyword-set*. The labeling process occurred in batches, progressively improving in agreement through iterative rounds. Initial disagreements notably stemmed from identifying sarcasm or instances where privacy violations were casually referenced while using these products as concerns (an example is given in Figure 6 in Appendix F). Ultimately, reviews with sarcastic tones expressing such concerns were included in our dataset. The use of sarcasm to convey these concerns finds support in existing research as well [22]. Despite disagreements, a consensus was reached to encompass a wide array of concerns, aiming to present a comprehensive portrayal of the identified issues in our analysis. In the labeling phase, the reviews that discuss the security and privacy aspects in a positive or neutral tone were excluded, assuming a portion of such reviews might be sponsored. Prior studies maintain that sponsored content infringes upon consumers' right to be informed when they are exposed to promotional material [19, 26, 46]. Further, negative reviews on S&P aspects might have more impact on a consumer's decision-making because of loss aversion, trust and credibility, and salience of negative information.

### 3.8 Thematic analysis of reviews with concerns from $\mathcal{D}_2$

A thematic analysis [24] was carried out on the subset of reviews within  $\mathcal{D}_2$  marked with S&P concerns (labeled as '1') to address RQ2. The inductive coding process followed guidelines outlined in the row "C. Thematic Analysis" in Table 4. Two coders were involved in it. At the initial meeting, the rules to be followed were discussed to make sure each coder followed the rules while labeling a review with a theme. The coding was independent, meaning a coder would not intervene with the other during the labeling step. It was also iterative, meaning the coders would sit for discussion after labeling a moderate number of reviews to discuss disagreements. Labeling followed by discussion and resolving disagreements was intended to reduce disagreements over the period and to obtain consistent labeling of themes. Figure 7 illustrates the process of labeling and resolving agreements. The overall inter-rater reliability (Cohen's  $\kappa$ ) was FT: 0.7437, SHS: 0.7709, and SSCS: 0.8576, which implies substantial, substantial, and almost perfect agreement, respectively [47]. Labeling *excerpts* was done in the order of FT, then SHS, and finally SSCS reviews, and the coders met to resolve agreements in between. We saw an interesting outcome of such an order in Cohen's  $\kappa$  since it gradually rose and reached its highest for the SSCS category. However, given there were discussions among the coders to resolve all conflicts the final labeling had full (100%) agreement. After labeling the *excerpts*, the low-level themes were grouped into high-level themes (See Figure 8).

### 3.9 Statistical tests

To address RQ3, we defined the *concern ratio* for each geographic region  $g$  as the ratio  $|C_g|/|R_g|$ , where  $|C_g|$  is the number of reviews representative of the *enhanced-keyword-set* with S&P concerns and  $|R_g|$  is the total number of collected reviews representative of the *enhanced-keyword-set* from region  $g$ . We used the Chi-squared test for *proportions* (R package *prop.test* [13]) and Fisher's exact test [1] when any cell counts in the contingency table were  $< 5$ . If the null hypothesis was rejected in these tests, we compared which pairs' concern ratios differed significantly from each other. During the posthoc test comparing multiple pairs, Bonferroni correction [3] was used to control for familywise error rate.

## 4 Device-type Specific Keywords

This section provides an overview of the findings related to **RQ1: How can we leverage consumers' IoT product reviews to identify S&P concerns?** We analyzed dataset  $\mathcal{D}_1$  to compile the *enhanced-keyword-set* for each device type. This facilitated the gathering of product reviews featuring these keywords, a substantial proportion of which voiced security and privacy concerns.

**Table 6.** The *enhanced-keyword-set* with their concern ratio for each device type (Row 1-3). The set of keywords from Fruchter et al. [36] (The bottom row).

Device Type	The <i>enhanced-keyword-set</i>	Concern-ratio
FT	card information, creepily, data privacy, data security, digital security, dubious, encrypted, goodbye privacy, hacked, hacker, hackers, intrusive, invade, invade privacy, invading, invasive, invasive of privacy, location services, personal data, privacy, privacy concerns, privacy invading, privacy-violating, requires location, security concern	29.58% (42/142)
SHS	beware of privacy, cautious, collecting data, hackers, illegal, intrusion, intrusive, invaded, invaded privacy, invades, invades personal privacy, invasion, invasive, listened 24/7, masquerading, network security concern, personal information, personal privacy, privacy, privacy concerns, privacy issues, privacy settings, recording, safety, securing data, security concern, security risk, spy , spying, threats , trojan,violation, wpa, wpa2	42.93% (792/1845)
SSCS	encrypts, extort, government, hacked, hackers, hacking, infiltrate, intrusive, moral code, password, permission, personal information, phone number, privacy, privacy issue, privacy mode, privacy policy, privacy reasons, privacy rights, sell data, snoop, spied, strange noises, violated, warrant	27.13% (239/881)
<b>Keyword chosen by Fruchter et al. [36] in their work</b>		
abuse, breach, confidential, creepy, crime, cyber, cybersecurity, damage, data, ethic, fbi, fear, firewall, freedom, government, hack, harm, individual, insurance, law, leak, legal, loss, license, agreement, malware, monitor, nsa, permission, physical, police, privacy, privacy policy, private, protection, secret, security, snowden, spy, spyware, storage, surveillance, terms of service, terrorism, third party, threat, track, unauthorized, violate, violation, virus, vulnerability, terrorism, violence		

#### 4.1 Effectiveness of the *enhanced-keyword-set* in finding reviews with S&P concerns


To assess the effectiveness of our *enhanced-keyword-set*, we computed the concern ratio, indicating the proportion of reviews within dataset  $\mathcal{D}_2$  expressing S&P concerns. Table 6 illustrates the concern ratio across each device type. The expertly curated examination of the subsampled dataset  $\mathcal{D}_1'$  significantly contributed to identifying a considerable volume of reviews highlighting S&P concerns through these keywords. Moreover, we noted variations in relevant keywords among different device types, reflecting distinct functionalities and contexts. Tailoring our keyword selection to each device type notably improved the likelihood of retrieving pertinent reviews compared to the usage of generic and over-broad S&P-related keywords in previous studies. For instance, when utilizing the keyword "law" as employed by Fruchter et al. [36], most of the extracted reviews were not related to security and privacy, such as discussion about "in-laws." Our approach enhanced the retrieval of pertinent reviews. For instance, FT-424, SHS-525, and SSCS-922 exemplify reviews obtained through type-specific keywords such as 'requires location,' 'recording,' and 'phone number,' respectively. These keywords emerged as the second most frequent for each device type in identifying concerns after the keyword 'privacy'. The full list of keywords used and their effectiveness in identifying reviews with concerns is available in Appendix E.

*FT-424 (searched with 'requires location'):* "This fitbit **REQUIRES location** services to be enabled to sync the fitbit with the android app...It won't function unless it can snatch your data even more than it already does."

*SHS-525 (searched with 'recording'):* "Like all voice assistance they **record** your voice. So privacy can and is a factor."

*SSCS-922 (searched with 'phone number'):* "They want your **phone number**. They want more and more personal information or you get locked out."

**Table 7.** The comparison of the effectiveness of the keywords of Fruchter et al. [36] vs. our *enhanced-keyword-set* in finding reviews with S&P concerns.

Domain	Device Type	Keyword set	No. of reviews with S&P concerns	No. of total reviews collected	Proportion of concern reviews (%)
	SHS	Fruchter et al.'s [36] keywords (Table 6, the bottom row)	691	4026	17.16
		Our <i>enhanced</i> keywords for SHS (Table 6, the second row)	529	1109	<b>47.70</b>

**Table 8.** The extent to which the two keyword sets (ours and the keywords from Fruchter et al. [36]) identify the same reviews

Type of reviews	No. of reviews	No. of concern reviews
Reviews found by our <i>enhanced-keyword-set</i>	1109	529
Common reviews found by both Fruchter et. al.'s [36] and ours	213	148
Row 1 - Row 2, i.e., reviews found only by our <i>enhanced-keyword-set</i> (%)	896 (896/1109 = 80.79%)	381 (381/529 = 72.02%)

Fruchter et al. [36]-562 (searched with 'law'): "We have had one for years. Got one recently for *mother-in-law*..."

#### 4.2 Comparison with keywords chosen by Fruchter et al. [36]

We conducted a comparison between the keyword set utilized by Fruchter et al. [36] and our *enhanced-keyword-set* for detecting S&P concerns in smart speaker reviews from the US Amazon store (Refer to Table 7). To make the comparison fair, we used the same set of products to collect reviews from and labeled the reviews collected with the search words from Fruchter et al. [36] with the same review labeling procedure as described in our methods. Additionally, since our reviews come from three domains and Fruchter et al.'s reviews come only from the US domain, we take the subset of SHS reviews in our dataset that comes from the US domain to make a fair one-to-one comparison. Using their keywords, we obtained 4,026 reviews, of which 17.16% (691/4,026) contained S&P concerns. Conversely, utilizing our *enhanced-keyword-set*, we obtained 1,109 reviews, with 47.70% (529/1,109) having S&P concerns. Table 8 shows that the intersection between the two keyword sets identifying the same reviews comprised of 213 (148 with concerns) reviews. Furthermore, we identified an additional 896 (381 with concerns) reviews that were not detected using Fruchter et al.'s keywords. This subset constituted approximately 80.79% (896/1109) of the total reviews and 72.02% (381/529) of the reviews with concerns found by our *enhanced-keyword-set*. Although with Fruchter et al.'s [36] keywords we were able to find more reviews with concerns (691 vs. 529), the proportion of S&P signals found with the *enhanced-keyword-set* was higher (17.16% vs 47.70%). A potential reason for this could be that our set of keywords is specific to each device type, while the keywords of Fruchter et al. [36] are over-broad and thus device-type agnostic.

**Takeaway:** Utilizing device-specific keywords and employing inductive coding, instead of relying on over-broad and generic keywords, allows for the extraction of reviews, a substantial proportion of which contain S&P concerns.

### 5 S&P Concerns about IoT Devices

This section presents the findings for **RQ2: What S&P concerns emerge from consumers' IoT product reviews?**

In line with §3.8, we identified twenty-one themes with overlaps (seven for each device type). After resolving the overlaps, there were fourteen distinct themes broadly falling into four categories: *data*-related, *technology*-focused, concerns about *surveillance*, and *superficial* concerns resulting from consumers' personal opinions. Table 9 illustrates the count of coded *excerpts* for each theme alongside their part-to-whole ratio.

**5.0.1 Data-related concerns.** This category includes themes related to data security and concerns about hackers. **Data security (FT, SHS, SCSS).** FT reviews included concerns about data collection, sharing contacts and personal information, and data breaches. This result further confirms the findings from Velykoivanenko et al. [68] who found that FT users are concerned about inference of their personality trait, socioeconomic status, political views, alcohol or tobacco consumption, or sexual activity from their fitness tracker devices and strengthen the conclusion drawn by Schomakers et al. [61] who found that users were even more worried about data privacy from fitness trackers as compared to remote monitoring of implanted cardiac devices.

*"They want to track my weight, food intake etc., hook up with their scale (separate purchase) charge me extra for an app, then share my information with others? Even my friends! The fitness industry is lame on this point."* (FT-395)

SHS reviews focused on concerns about data collection, retention, and the need for secure data storage within the Alexa ecosystem. Users expressed skepticism about sharing sensitive data, raised concerns about data encryption, and about sharing account details for payments.

*"Your personal data is collected without your permission from a device for which you paid for...scary isn't it."* (SHS-3164)

In SCSS, data-related concerns were the most frequent. Users expressed concerns about sharing phone numbers to use the app and were skeptical about sharing login information with other users. Users were also concerned about the lack of local storage for surveillance data and preferred not to store sensitive information in the cloud. Other common concerns included sharing personal information, general comments on data security, and data collection.

*"PrivacyBlink demands that you provide your phone number. Why? There is no conceivable reason why Blink would need your phone number, and it's just one more thing they will misuse or accidentally give away in a data breach. Overall, a VERY poor wireless camera."* (SCSS-1030)

**Table 9.** S&P concerns about IoT devices. A dash ('-') implies that the theme was not found for that device type.

Category	Theme	No. of excerpts coded (part-to-whole ratio)			
		FT	SHS	SCSS	Aggregate
Data-related	Data security	7 (0.11)	119 (0.12)	<b>149 (0.46)</b>	275 (0.1937)
	Concerns or vignettes about hackers	-	13 (0.01)	47 (0.15)	60 (0.0423)
Technological	Wi-Fi security	2 (0.03)	-	-	2 (0.0014)
	Payment security	7 (0.11)	-	-	7 (0.0049)
	Tracking	16 (0.25)	-	-	16 (0.0113)
	Lack of trust in the device	-	111 (0.11)	8 (0.02)	119 (0.0838)
	Issues with apps	<b>20 (0.31)</b>	-	66 (0.20)	86 (0.0606)
	Technical and technological issues	-	88 (0.09)	-	88 (0.062)
	Invasive customer support	1 (0.02)	-	-	1 (0.0007)
Surveillance	Surveillance by a government or a regime	-	38 (0.04)	3 (0.01)	41 (0.0289)
	Affiliation with a country	-	-	21 (0.07)	21 (0.0148)
	Company's affiliation	11 (0.17)	-	-	11 (0.0077)
Superficial	Intrusion or invasion	-	<b>457 (0.44)</b>	-	<b>457 (0.3218)</b>
	Personal skepticism and opinions	-	207 (0.2)	29 (0.09)	276 (0.1662)
<b>Total</b>		<b>64 (1.00)</b>	<b>1033 (1.00)</b>	<b>323 (1.00)</b>	<b>1420 (1.0000)</b>



**Concerns or vignettes about hackers (SHS, SSCS).** SHS had limited *excerpts*. Users mainly expressed concerns about hackers and shared personal experiences of hacking. Some of the concerns we uncovered for SHS show skepticism about the device being hacked:

*"Could it be used for a listening device for certain agencies or hackers? I would have to say yes, so I have to watch what I say around her. LOL" (SHS-1207)*

For SSCS, this theme was the third most prevalent one. Users shared experiences of being hacked and expressed concerns about hacking incidents, sharing specific vignettes:

*"Also someone has been trying to hack mine, I'm constantly having to change my passwords and emails, it's very uncomfortable!" (SSCS-727)*

**5.0.2 Technological concerns.** In this section, we discuss themes centered around different aspects of technology lacking required security or privacy, including issues with the device and support ecosystem.

**Wi-Fi security (FT).** FT users complained about the app's inability to accept special characters in Wi-Fi passwords and its constant requirement for a connection:

*"Found out after purchasing that it needs to connect to wifi for all the features to work. It also can not accept special characters in the wifi password. So you either have to live with only bluetooth connectivity and no music, or dumb down the security on your wifi. Personally, I will not compromise my digital security for Fitbit's inability to design their software." (FT-380)*

**Payment security (FT).** In this theme, we highlight payment security concerns, particularly regarding the sharing of credit card information. Users raised safety concerns about the device's contactless payment feature. Some mentioned the app's lack of an independent payment system, while others expressed reluctance to use the device for payments.

*"Like for the pay thing, there's one more way to have a credit card number out-there, so when FitBit is hacked, there's another cyber thief to get such information." (FT-257)*

**Tracking (FT).** FT users expressed concern about location sharing, as some apps required it for syncing. Reviews also mentioned constant prompts to enable location services, raising privacy concerns. Some users feared constant monitoring by the product.

*"Shortly after setting up my Fitbit tracker the company pushed a software upgrade to the app that works with it. Now, every time I open the app I get a prompt asking me to turn on location services. Well, I had the app set to allow location access only when using, but apparently that's not good enough. Fitbit wants to track my every movement!" (FT-524)*

**Lack of trust in the device (SHS, SSCS).** Some SHS concerns focused on the device's microphone and the option to unplug it. Some users specified unplugging the device for privacy, while others discussed disabling or muting the microphone. Users also expressed skepticism about placing the device in private areas.

*"I unplug it when not in use. That takes care of the recording and camera." (SHS-2990)*

For SSCS, users expressed distrust in using these devices, but the number of concerns in this category was minimal. Half of the related reviews discussed unplugging the device for privacy, while the remaining reviews mentioned turning off the microphone or using the physical camera shutter for privacy.

*"We, mostly, leave them unplugged (for privacy) but use them whenever we're away and have a dog sitter or when the cleaning crew comes." (SSCS-546)*

**Issues with apps (FT, SSCS).** FT users raised concerns about account hacking and expressed skepticism about granting permissions. They also had issues regarding apps running constantly, affiliation with third-party apps, integration with Alexa, and the presence of hack bots. Additionally, there were mentions of invasive apps, excessive data requests, the lack of Multi-Factor Authentication (MFA), and receiving unsolicited text messages.

*"...despite this being a device that contains your personal health information, you cannot protect your fitbit account with multifactor authentication." (FT-239)*

Within SSCS, users were skeptical about granting permissions for personal information. They also expressed worry about the privacy of their audio data. Additionally, they considered the security aspects of their routers, Wi-Fi, and the overall security of their networks. Specifically, they were concerned about changing passwords

without special characters. Additional considerations included authentication and password security, lack of encryption, issues with third-party apps, and receiving spam. For instance, a user expressed concerns as to why an app needed all types of permissions and described manually disabling most of them:

*"Home app requires almost all the permissions from your phone. But, I was able to disable most permissions except location and storage." (SSCS-1440)*

**Technical and technological issues (SHS).** Some of the users' concerns arose from technical aspects impacting S&P. The most commonly observed instance was that of Alexa activating without wake words. Another aspect involved security concerns with the Drop-In feature [9]. Users also expressed concerns about the Amazon Sidewalk feature and privacy over voice commands. Other minor themes included app privacy features, wifi, and password security.

*"Also I dislike the fact that this Item Still Records Voice Recognition without the Awake word being Commanded. Privacy Intrusion." (SHS-3001)*

**Invasive customer support (FT).** Some users of FT also expressed concerns about invasive behavior by customer support:

*"If you have any type of wifi that uses real security, like MAC filtering or network profiling don't bother contacting support. They can creepily tell you that your versa2 is connected to your phone and what the battery charge is, (probably more) but that's it." (FT-374)*

**5.0.3 Concerns about surveillance.** Here, we highlight concerns that stem from potential affiliations and associations with entities, raising concerns about some form of consumer surveillance.

**Surveillance by a government or a regime (SHS, SSCS).** SHS users discussed concerns about government surveillance, referring to a totalitarian regime resembling Big Brother [56]. Users believed that Alexa might be sharing information with the government.

*"It's ok for a government spy." (SHS-2270)*

Similar to SHS, reviews for SSCS also mentioned the theme of government surveillance, referencing Big Brother [56].

*"I can no longer access my cameras without giving my cell phone number to Amazon. Wtf? I used to like Blink. You know what I don't like? Big Brother." (SSCS-956)*

**Affiliation with a country (SSCS).** Most SSCS devices we chose had manufacturing or support servers from a specific region, with their reviews raising significant privacy and security concerns. The most common observation was distrust in the software from that region. Some user concerns were also about data traffic and data influx into servers located in that region.

*"All good but privacy may concern as it is Chinese product and its server location at China." (SSCS-1458)*

**Company's affiliation (FT).** FT users expressed concerns about the acquisition of small companies by the big ones and the potential sharing of their data. They were skeptical about using an account from the acquiring company due to security concerns. This confirms the finding from Lidynia et al. [50], Becker et al. [21], and Vitak et al. [70] works who also found that FT users are highly concerned about possible recipients of their private data, sharing of PHI (Protected Health Information) with third-parties, and selling their personal data.

*"Fitbit company was acquired by Google shortly after I purchased this. Cannot use Fitbit syncing now without enabling every form of dozens of different kinds of privacy-violating and tracking services with Google." (FT-525)*

**5.0.4 Superficial concerns.** Here we show consumers' potential misconceptions, highlighting how these can stem from an incomplete mental model and information asymmetry, while others from lack of technical knowledge and conviction.

**Intrusion or invasion (SHS).** For SHS devices, the highest number of concerns obtained from our review corpus were related to device recording and listening to conversations. Users expressed discomfort with the device's intrusive nature and receiving targeted ads based on eavesdropping. These reviews generally lacked specific incidents or anecdotes.

*"One thing that is pretty weird is that after I got her I got ads about how alexa is recording and saving the audio and sending it to amazon to advertise to us." (SHS-1932)*

**Personal skepticism and opinions (SHS, SSCS).** This theme emerged as the second most prominent for SHS users after the "intrusion or invasion" theme. Users expressed skepticism about the product's privacy aspects. However, we also observed reviews where users mentioned accepting privacy violations while still using the device for its functionality. We note that consumers tend to use sarcasm to express S&P concerns.

*"It has great potential to invade privacy, and I have put it away, never to be used." (SHS-1613)*

*"Well it's a room spy and it kinda does what is says on the tin so. Want a room spy then here we have one lol." (SHS-776)*

For SSCS, the *excerpts* in this category expressed skepticism about the device's privacy and security aspects. Two reviews mentioned technical terms like "NAS" (Network Attached Storage) and "WPA" (Wifi Protected Access) in relation to S&P, revealing that consumers with a wide range of S&P knowledge contribute to these reviews, further emphasizing the merit in considering product reviews as a good source of understanding prevalent S&P concerns in consumers.

*"For a security product, it is expected that the account management happens according to the latest standards and practices. The password is limited to 32 characters and does not allow for spaces. According to NIST 800-63B Digital Identity Guidelines, 5.1.1.2 Memorized Secret Verifiers: "Verifiers SHOULD permit subscriber-chosen memorized secrets at least 64 characters in length. All printing ASCII [RFC 20] characters as well as the space character SHOULD be acceptable in memorized secrets. Unicode [ISO/ISC 10646] characters SHOULD be accepted as well." (SSCS-1365)*

**Takeaway:** The major portion of the S&P concerns are superficial. They may stem from users' incomplete mental models and information asymmetry between buyers and sellers. Data security is a concern across all device types. Technological themes mostly cover issues with apps (FT and SSCS) and users' trust in the device (SHS). The rest of the themes emerge from users' concerns about being surveilled by any company or regime.

## 6 S&P concerns across geographic regions

This section addresses **RQ3: How do IoT consumers' S&P concerns vary across geographic locations?** We collected product reviews from `amazon.com`, `amazon.co.uk`, and `amazon.in` representing the US, the UK, and India, respectively. We explore if the *concern ratio* (introduced in §3.9) significantly differs across regions. Note that the reviews collected in our study represent the reviews containing any words from the *enhanced-keyword-set*. Therefore, the answer to **RQ3** is a statistical inference from a sample that represents the reviews containing those enhanced keywords (as stated before in §3.6).

### 6.1 Statistical tests to compare concern ratios across geographic regions

To compare the *concern ratios*  $|C_g|/|R_g|$  among regions (as introduced in §3.9), we conducted Chi-squared tests, except for FT, where Fisher's Exact Test was employed because a cell count in the contingency table was  $< 5$ . In Table 10, the middle column displays the test outcomes, examining the null hypothesis  $H_0 : |C_g|/|R_g|$  is the same across all regions  $g$ , while the alternative  $H_A : \text{At least one } |C_g|/|R_g| \text{ for a region } g \text{ differs from the others}$ . We rejected  $H_0$  for all three device types at a significance level of  $\alpha = 0.05$ . This implies that for all device types, at least one region's concern ratio differs significantly from others. Since  $H_0$  was rejected for all device types, we proceeded with a posthoc pairwise comparison test of  $|C_g|/|R_g|$  across all pairs of geographic regions  $g$  within each device type. This investigation aimed to identify pairs whose concern ratios significantly differ from each other. The results in Table 10, displayed in the rightmost column, suggest that for all device types, the concern ratio in the US significantly differs from that in the UK. Additionally, for SSCS, the concern ratio in the US significantly differs from that in India. Across all the US vs. the UK pairs, the concern level in the US is significantly higher than in the UK. Similarly, for SSCS, the concern ratio in the US is significantly higher

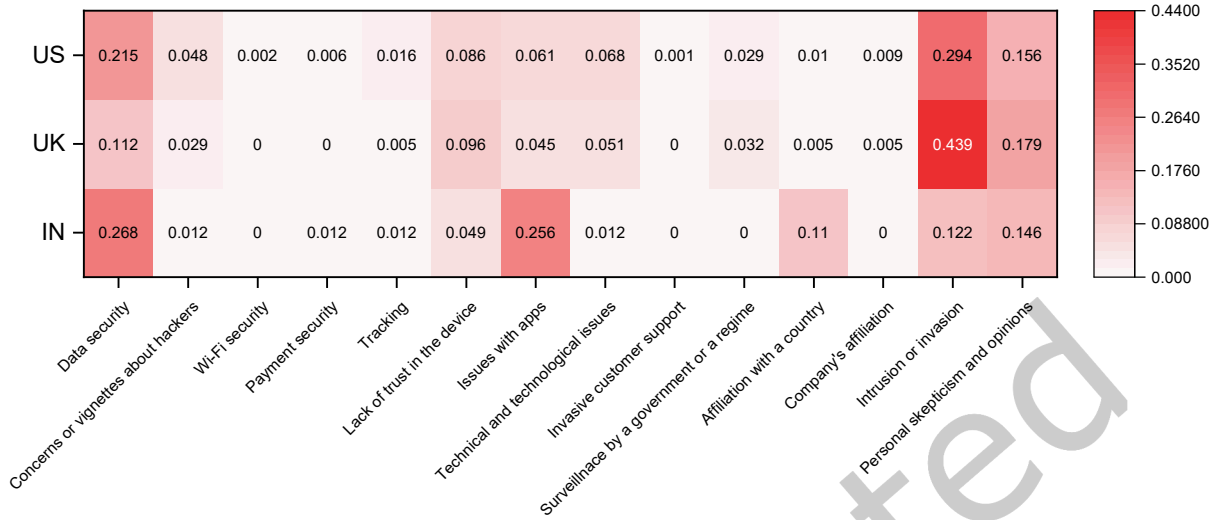
than in India (refer to the 'Concern ratio' column in Table 10). We observe that for the category FT and SHS, the Indian domain has the highest concern ratio. Therefore, in the posthoc pairwise tests for those categories, it is counterintuitive to notice that the Indian domain, at a significance level of  $\alpha = 0.5$ , failed to reject the hypothesis that the concern ratio of the Indian domain is not significantly different from that of the US and UK domains. This is because the number of reviews found in the Indian domains was much less compared to the other two domains. For FT, the choice of Fisher's Exact test instead of the Chi-squared test for proportion took care when the count in a cell of the contingency table was  $< 5$ . For SHS, none of the cell counts in the contingency table was  $< 5$ . Nevertheless, the concern ratio of the Indian domain was not significantly different from that of the other two domains because of the moderately small number of reviews even though the sample size was statistically large, i.e.  $n > 30$ .

## 6.2 Themes across geographic regions

Here, we discuss the distribution of the themes across geographic regions. Figure 4 shows the heatmap of geographic region vs. the fourteen themes that emerged with the thematic analysis. In this heatmap, for a cell at row  $i$  and column  $j$ , the cell value denotes the ratio of the number of excerpts coded with theme  $j$  to the total number of excerpts coded from geographic region  $i$ . Therefore, it is a part-to-whole ratio =  $|E_{theme}|/|E_{region}|$  for excerpts where all values in a row sum up to 1. For the rest of the section, we will denote this part-to-whole ratio for excerpts as  $|E_{theme}|/|E_{region}|$ . If some cells for any region had zeros in them, it means from that region, we found no excerpts that can be coded with that theme. In total, 1234 excerpts were coded where 845 were coded from the US domain, 325 from the UK domain and 64 from the Indian domain. Hence  $|E_{region=US}| = 845$ ,  $|E_{region=UK}| = 325$ , and  $|E_{region=IN}| = 64$ . In the US and the UK, the value for  $|E_{theme}|/|E_{region}|$  were the highest for the theme 'Intrusion and Invasion' (US: 0.294, UK: 0.439). For India, it was for the theme 'Data security' (0.268). In the US (top row), we found at least some excerpts coded with all the themes. In the UK, no excerpts were coded with the themes 'Invasive customer support', 'Payment security', and 'Wi-Fi security'. For India, we found no excerpts coded with 'Invasive customer support', 'Company's affiliation', 'Wi-Fi security', and 'Surveillance by a government or a regime'. Such findings give us an idea of what themes the users from each region are concerned or not concerned about. Obviously, such an observation is limited by the sample of reviews gathered

**Table 10.** Calculation of the concern ratio (left). Test for equality of proportions.  $H_0 : |C_g|/|R_g|$  is the same in each  $g$ .  $H_A$  : At least one  $|C_g|/|R_g|$  for a  $g$  is different from the others (middle). Posthoc pairwise comparison of the proportions of all pairs of  $|C_g|/|R_g|$ -s (right). The significance codes \*\*\*, \*\*, and \* mean  $p < .001$ ,  $p < .01$ , and  $p < .05$ , respectively.

Type	Concern ratio calculation				Test for equality of proportion				Pairwise comparison		
	Geographic region $g$	No. of reviews with S&P concerns $ C_g $	Total no. of reviews $ R_g $	Concern ratio $ C_g / R_g $	Type of test	Test statistic	df	p-value	$g_i$	$g_j$	p-value
FT		33	100	0.33	Fisher's exact	-	-	0.01049*			0.020*
		5	33	0.1515							1.0
		4	9	0.4444							0.084
SHS		529	1109	0.477	Chi-squared	$\chi^2=30.675$	2	2.182e-07***			2.3e-07***
		239	689	0.3469							1.0
		24	47	0.5106							0.1
SSCS		181	582	0.311	Chi-squared	$\chi^2=13.778$	2	0.001019**			0.024*
		33	164	0.2012							0.015*
		25	135	0.1852							1.0



**Fig. 4.** The heatmap for geographic region vs. theme. Each row denotes a geographic region  $g$ . Each column denotes a main theme from the thematic analysis. The entry in each cell  $i, j$  of the heatmap denotes the part-to-whole ratio =  $|E_{theme=j}|/|E_{region=i}|$  of the number of excerpts coded with that theme  $j$ ,  $|E_{theme=j}|$  to the total number of excerpts (highlighted parts in a review) coded in geographic region  $i$ ,  $|E_{region=i}|$ . An excerpt is a part of a review that is highlighted with security and privacy concerns. All values in a row sum up to 1.

with the *enhanced keyword set*. If we observe the heatmap column-wise, we notice that ‘Data security’, ‘Intrusion or invasion’, and ‘Personal Skepticism and opinions’ are prevalent across all three regions. Another thing we notice from the heatmap is that the theme ‘Issues with apps’ and ‘Affiliation with a country’ has an alarmingly high  $|E_{theme}|/|E_{region}|$  value in India compared to others. This was because the users in India are more concerned about the security and privacy of the companion apps of the product they are using. Further, they might be concerned that a product that is manufactured or has servers at a certain region on earth may affect their data security. Our thematic analysis of SSCS excerpts shows that many users are worried about such a concern. We also notice a contrast between India and the other two regions for the following themes: ‘Lack of trust in the device’, ‘Surveillance by a government or regime’, and ‘Issues with technical and technological aspects’. This may imply that the US and UK users are more concerned about these themes than users from India.

**Takeaway:** Reviews from the US domain express a significantly higher proportion of security and privacy concerns than the UK domain for all device types and the Indian domain for security cameras.

## 7 Discussion

### 7.1 Experimental design choices

**Steps to encounter fewer false positives and false negatives.** The job of labeling the reviews with 0-s (no S&P concerns) and 1-s (S&P concerns) in the review labeling step by two coders worked as a check and balance to encounter false positives and negatives. The primary coder, along with the secondary coder met at regular intervals to keep up with the consistency of the labeling. False positives and negatives did arise. We state an example of a false negative (a true 1 labeled initially as a 0) at first. It is the use of sarcasm to express concerns that were found by the second coder. Therefore, reviews with such a tone were later labeled accordingly by both of them (previously 0, later as a 1). False positives are examples of human errors (labeling a 0 as a 1). Since two coders mutually labeled a review, the likelihood of both labeling a review incorrectly is potentially very low.

We quantify false positives (FP) or false negatives (FN) as instances where exactly one coder mislabeled a row, and the label of the other coder was used to label that row correctly. The counts are as follows: FT: 25/142 (FP: 11/100, FN: 14/42), SHS: 280/1845 (FP: 146/1053, FN: 134/792), and SSCS: 107/881 (FP: 34/642, FN: 73/239). The other example of false negatives is the reviews that did not contain the enhanced keywords but did have concerns. While this is a limitation of our search procedure, we leave as future work to generalize a Machine Learning model (e.g., fine-tuned large language models) that can detect security and privacy concerns even if they do not contain the enhanced keywords from this human-driven study.

**Not taking a union of three sets of keywords.** We avoided combining three sets of keywords for several reasons. The first reason stems from our experimental design. Both the coders agreed to choose keywords they found by reading a review related to S&P concerns. This makes the choice specific to the review and the context described in that review. We hypothesize the context described in a review is related to the functionality or category of the device. Hence, keywords eligible for being in the *enhanced-keyword-set* for one category might be considered noisy in another category. Any overlap between the keywords across categories is purely coincidental since we do not choose a keyword for a category because it was present in the reviews for another category. This would contradict our design. For FT, an example of a noisy keyword is 'security'. When searching with this keyword in FT, we obtained a lot of noisy reviews that talk about the strap of the tracker not securely being placed on the wrist.

*(Searched with 'security', Fitbit Charge 4, ASIN: B084CQ41M2): "Bought watch in May. Last week (July) stopped getting notifications from text or phone. Today, still July, watch face will no longer hold band securely on one side. Don't know who to notify as return date is past."*

For other device categories, the term "security" alone did not qualify for inclusion in the enhanced keyword set, indicating its noise level. It was either combined with "concern" (e.g., "security concern" for SHS and SSCS), with "risk" (e.g., "security risk" for SSCS), or transformed into a different form (e.g., "securing data" for SSCS) before being added to the enhanced keyword set.

## 7.2 Implications and Recommendations

We grouped the fourteen themes of concerns into broader categories, as shown in Table 9. Based on these data-related, technological, surveillance-based and superficial concerns expressed by consumers, we offer some recommendations for IoT ecosystem stakeholders.

**Clear Statement of Data Collection Policy.** Concerns related to data security and hacking highlight the lack of consumer awareness regarding data collection and its purpose [37]. Product manufacturers can alleviate misplaced concerns by offering concise and transparent statements delineating their data collection policies, purposes, and retention periods. This step is crucial given that users often overlook privacy policies and terms, leading to misconceptions. Incomplete mental models of IoT products contribute to errors and skepticism about S&P implications [16, 48]. Collaboration among app developers, manufacturers, and policymakers is vital to address these concerns. Valid concerns will help manufacturers improve their products, while misinformation in consumer reviews will enable them to include authentic information in their product web pages or online forums so that consumers will not refrain from using these products due to the spread of misinformation.

**Surveillance by Government or Regime.** Users may develop undue concerns about surveillance by a government or a regime, potential affiliations with foreign entities, or specific companies if they do not review the product's privacy policies. Consumers often overlook privacy policies, terms, and agreements when making purchases [37]. To address this, policymakers should clearly disclose any data sharing with governments or corporations, providing explanations for such affiliations. This transparency would help users avoid misinformation about regime surveillance or corporate acquisitions. Some concerns in our thematic analysis appear superficial and mainly consist of users expressing skepticism about product's S&P without providing substantial details or insights. Although lacking depth, these themes are still significant. Information asymmetry, where

manufacturers possess more knowledge than customers, contributes to the skepticism surrounding products that collect sensitive data. Consumers find the process of obtaining relevant S&P-related information challenging [32]. This information gap may lead users to express concerns vaguely and superficially.

**The Role of E-commerce Platforms.** E-commerce platforms that have IoT product listings have an important role to play in influencing consumers' purchase decisions and reducing skepticism. Platforms such as Amazon provide several categories of reviews relevant to the product (e.g., "Performance", "Ease of setup", "Sound quality" for Echo Show 8 [20]) as part of the user interface. Including "security and privacy" as one such category can enable S&P-focused consumers to make better purchase decisions. Sub-categories related to "S&P" reviews in line with the types of concerns we have uncovered may also provide more fine-grained control to consumers. In addition to highlighting negative S&P reviews, the platform must also highlight positive S&P-related reviews to aid in decision-making.

In addition, e-commerce platforms can also enable vendors and manufacturers who have product listings to view product reviews sorted by categories of reviews, thus facilitating them to focus on specific concerns expressed by consumers. Platforms can additionally validate the veracity of information in the reviews and the correctness of the review categorization through crowd-sourcing. Speculative or superficial concerns may be flagged by manufacturers or e-commerce platform users with high reputations. Manufacturers can instead provide relevant documentation that refutes speculative concerns, thus increasing the reliability of S&P reviews about a product.

**Cross-geographic Implications.** Significant concern ratio differences were observed between the UK and the US for all device types (§6). The UK's implementation of GDPR, including the Data Protection Act of 2018 [7] and the UK GDPR [8] introduced in 2021, reflects stricter regulations compared to the US counterparts like CCPA [5]. These regulations establish comprehensive principles, rights, and obligations for personal data processing. Similarly, the concern ratio in India for SSCS was significantly lower than that of the US, potentially due to cultural differences [59]. Manufacturers should consider comparing and adopting policies from other regions to address significant security concerns of their consumers to comply with regional regulations and cultural differences. While we did not observe any explicit mentions of the GDPR or CCPA in the consumer reviews, we did see consumers criticizing Amazon's policy in SHS reviews and skepticism around security cameras having their servers located in a certain country. Such reviews imply that people may think certain companies collect data and share it with the government, and the location of servers in a certain region on Earth can affect their data security. Following are some reviews to support such insights.

*SHS-3095: "...Now with all our local neighborhood Amazon devices connecting and monitoring every activity from me, my neighbors, and any nearby stranger I refuse to partake in BIG BROTHER AMAZON COLLECTING, SELLING OR OTHER WISE OWNING MY EVERY SPOKEN WORD, ON SCREEN CAPTURED VIDEOS / PHOTOS, CORRELATED TO MY EVERY STEP VUA GPS /TRIANGULATION, AND THE DEVICES I USE. Really Amazon.... It appears you have overstepped your rights and responsibilities regarding collecting, selling, and giving stare federal agencies user's private information...."*

*SSCS-1363: "Anyone paying cheap monthly fee for a cloud service should consider that his recordings may be stored abroad (Asia-China) and therefore abroad the privacy laws are not the same than in EU."*

**Relation between Product Ratings and S&P Concerns.** Besides answering the three research questions, we conducted an additional analysis comparing the average ratings between concern and non-concern reviews. Users expressing concerns tend to assign lower star ratings. The *t*-test (Table 11), supported by Levene's test for equality in variance [2] (Table 12), shows that concern reviews have significantly lower average star ratings than non-concern reviews. This indicates that dissatisfaction with S&P aspects can impact user ratings. Understanding online ratings aids retailers in demand forecasting [34]. Our recommendation for IoT product retailers is to scrutinize lower-rated reviews to determine if dissatisfaction is solely due to security and privacy issues. If so, enhancing these aspects could improve customer satisfaction and ratings.



**Table 11.** 2-sample  $t$  test.  $H_0 : \mu_c = \mu_n$ , vs.  $H_A : \mu_c < \mu_n$ .  $\bar{y}_c$  and  $\bar{y}_n$  are the sample average ratings of the concern and non-concern reviews, respectively. The significance code \*\*\*, \*\*, and \* means  $p < .001$ ,  $p < .01$ , and  $p < .05$ , respectively.  $n$  refers to sample size.

Device type	$\bar{y}_c$	$\bar{y}_n$	$t$	$df$	$p$ - value
FT	2.24 ( $n = 42$ )	3.17 ( $n = 100$ )	-3.511	140	0.0003***
SHS	3.25 ( $n = 792$ )	4.37 ( $n = 1053$ )	-16.224	1407.1	2.2e-16***
SSCS	1.90 ( $n = 239$ )	2.68 ( $n = 642$ )	-7.2522	556.56	6.9e-13***

**Table 12.** Levene's test for equality in variances.  $H_0 : \sigma_c^2 = \sigma_n^2$  vs  $H_A : \sigma_c^2 \neq \sigma_n^2$ .  $s_c^2$  and  $s_n^2$  are the sample variances in the ratings of the concern and non-concern reviews, respectively. The significance code \*\*\*, \*\*, and \* means  $p < .001$ ,  $p < .01$ , and  $p < .05$ , respectively.

Device type	$s_c^2$	$s_n^2$	$k$	$n$	$F$ distribution	$F$ - value	$Pr(> F)$
FT	1.34	1.47	2	142	$F_{1,140}$	0.7598	0.3849
SHS	1.62	1.21	2	1845	$F_{1,1843}$	242.08	2.2e-16***
SSCS	1.28	1.68	2	881	$F_{1,879}$	107.56	2.2e-16***

In summary, our review analysis reveals new findings not explored in previous research. To the best of our knowledge, we are the first to conduct a cross-geographic analysis on popular device categories. Our carefully curated category-specific keywords effectively eliminate noise and account for language variations, enabling us to focus on a product corpus where a substantial number of S&P concerns are discovered.

### 7.3 Limitations

**Only one manufacturer for the Fitness Tracker and Smart Home Speaker category.** To some extent, the generalizability and applicability become limited in terms of the variability of manufacturers since our products from FT and SHS come from only one manufacturer. However, since human labor is the essence of our work, we had to limit ourselves to reading a moderately large dataset that is not too large to cause human error and fatigue in their analysis. This is why we chose the top three products based on their global ratings. Empirical evidence showed that products with more global ratings have more reviews posted by users, and products with more reviews increase the likelihood of obtaining reviews with certain keywords. Although not included, we did go into the webpages of some less popular products that have a modest amount of reviews to see if we found some reviews with certain keywords. Even with a keyword like 'privacy' when we found no results, we decided to focus on the popular ones.

**Seeding Search Keywords.** Relying solely on the keyword 'security' and 'privacy' in  $\mathcal{D}_1$  may have resulted in the exclusion of other product reviews. We acknowledge this as a limitation of our study. While we agree that the reviews filtered by our approach may be biased toward users who may have a S&P-informed mental model, it is more likely this population will provide us with different nuanced S&P concerns than the general population. According to the second bullet point of Table 4, we choose words from  $\mathcal{D}'_1$  by reading only the reviews that voice security and privacy concerns. Therefore, our chosen words are very specific to the concerns in the reviews that we found and read. We hypothesize a person's choice of words is directly related to what they want to express. Since they want to express some concerns related to security and privacy, we focus on those words. The overlap of the keywords between categories can happen coincidentally. Our goal was to target a specific niche in the product corpus where significant security and privacy concerns would prevail. The focus on

**Table 13.** Time-related information of  $\mathcal{D}_2$ 

Device type	All reviews in $\mathcal{D}_2$		Reviews with concerns in $\mathcal{D}_2$	
	From	To	From	To
FT	2019-09-16	2022-02-03	2019-09-16	2022-01-27
SHS	2018-10-16	2022-02-02	2018-10-20	2022-01-31
SSCS	2018-11-18	2022-02-03	2018-11-18	2022-02-03

‘security’ and ‘privacy’ as initial keywords unsurprisingly revealed *data security* as a primary concern, aligning with our intention to identify a niche with substantial S&P concerns in the review corpus. Our intuition to adopt a restrictive approach proved successful, as our findings confirmed the presence of well-founded S&P concerns in the identified niche. Although our sample showed bias towards products from specific vendors in the initial qualitative phase, it provided valuable insights from popular products with numerous consumer reviews. While the sample bias limits generalizability, we consider it a necessary first step in understanding user concerns. In future research, we aim to explore more diverse products and vendors to comprehend potential variations in raised concerns.

**Temporal Evolution of Reviews.** While it is reasonable to infer security and privacy concerns from product reviews, it is important to note that the time when a user posts a review may not necessarily align with their actual perceived security and privacy concern. The dataset  $\mathcal{D}_2$  collected with the enhanced keyword set consists of reviews posted between late 2018 to early 2022 (Table 13 lists the review timeline). In  $\mathcal{D}_2$ , the oldest review is from October 16th, 2018 and the newest review is from February 3rd, 2022. In the distant future, some concerns in  $\mathcal{D}_2$  might have already been resolved by manufacturers, while users who did not post concerning reviews might discover unaddressed concerns.

**Broad Definition of S&P.** Moreover, while we started with a comprehensive definition of S&P concerns, we encountered vagueness and anecdotal content in many reviews. This constrained our effort to map each review to all the aspects defined in the S&P definition. While labeling the reviews, we still kept those additional aspects of confidentiality, integrity, authentication, secrecy, anonymity, and solitude in our mind and made an implicit mapping. Table 17 of Appendix F lists examples of reviews mapping different aspects of the S&P definition.

**Lack of Large Number of Samples.** Another limitation is the limited availability of a large sample of reviews for the FT device type. Additionally, the human labor involved in our methodology makes it impractical for very large-scale integration. Still, our findings provide valuable insights about the types of concerns that consumers have and what factors motivate those concerns. In the future, we plan to extend our work by automating different steps, enabling us to analyze a larger set of devices and reviews. We have developed a labeled dataset containing 1,378 *excerpts*, which we plan to release to the community so it may serve as a benchmark for identifying IoT security and privacy concerns. We intend to explore using large-language models and text entailment techniques [42] to automatically identify similar reviews and expand the dataset.

**Language in reviews.** Part of our preprocessing includes filtering out any reviews written in languages other than English (see §3.2). This might cause some reduction in data for domains where English is not predominantly spoken, i.e., the Indian domain. In the future, we plan to apply automation for this human-driven work. Part of this automation would include translating reviews from other languages into English and then considering them for classification and thematic analysis related to security and privacy concerns.

## 8 Conclusion

In this paper, we extract security and privacy concerns from IoT device reviews, focusing on fitness trackers, smart home speakers, and security cameras. Analyzing Amazon reviews in the US, UK, and India, we identified

fourteen themes grouped into four categories. Cross-geographic analysis revealed higher concern ratios in the US than in the UK for all device types. Our study demonstrates leveraging online reviews for addressing security and privacy risks, enhancing products, educating users, and fostering trust in IoT adoption.

## References

- [1] 1954. Fisher's exact test. <https://mathworld.wolfram.com/FishersExactTest.html>.
- [2] 1960. Levene test for equality of variances. <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>.
- [3] 1961. Bonferroni correction. <https://mathworld.wolfram.com/BonferroniCorrection.html>.
- [4] 2017. Worldwide Wearables Market Grows 7.3% in Q3 2017 as Smart Wearables Rise and Basic Wearables Decline, Says IDC. <https://www.businesswire.com/news/home/20171130005146/en/Worldwide-Wearables-Market-Grows-7.3-in-Q3-2017-as-Smart-Wearables-Rise-and-Basic-Wearables-Decline-Says-IDC>. Accessed: 2022-09-13.
- [5] 2018. The California Consumer Privacy Act. <https://theccpa.org/>
- [6] 2019. *Consumer Internet of Things Security Labelling Survey Research Findings*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/950429/Harris\\_Interactive\\_Consumer\\_IoT\\_Security\\_Labelling\\_Survey\\_Report\\_V2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950429/Harris_Interactive_Consumer_IoT_Security_Labelling_Survey_Report_V2.pdf)
- [7] 2021. UK to overhaul privacy rules in post-Brexit departure from GDPR. <https://www.theguardian.com/technology/2021/aug/26/uk-to-overhaul-privacy-rules-in-post-brexit-departure-from-gdpr>. Accessed: 2022-09-14.
- [8] 2022. Guide to the UK General Data Protection Regulation (UK GDPR). <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/>. Accessed: 2022-09-14.
- [9] 2022. How Does Drop In Work? <https://www.amazon.com/gp/help/customer/display.html?nodeId=GS3WRTSRKD2U6MCK>. Accessed: 2022-09-11.
- [10] 2022. State of IoT 2022: Number of connected IoT devices growing 18% to 14.4 billion globally. <https://iot-analytics.com/number-connected-iot-devices/>. Accessed: 2022-09-11.
- [11] 2023. Do fitness trackers put your privacy at risk? <https://usa.kaspersky.com/resource-center/preemptive-safety/fitness-tracker-privacy>.
- [12] 2023. Global ratings on Amazon reviews. <https://www.ecomengine.com/get-amazon-reviews>. Accessed: 2023-04-8.
- [13] 2023. prop.test: Test of Equal or Given Proportions. <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/prop.test>. Accessed: 2023-01-13.
- [14] 2023. Simple random sample. [https://en.wikipedia.org/wiki/Simple\\_random\\_sample](https://en.wikipedia.org/wiki/Simple_random_sample). Accessed: 2023-11-21.
- [15] 2024. Amazon Product Ratings and Average Ranks | What They Hide and How to Improve Them. <https://feedcheck.co/blog/amazon-product-ratings-and-average-ranks-what-they-hide-and-how-to-improve-them/>. Accessed: 2024-05-17.
- [16] Noura Abdi, Kopo M Ramokapane, and Jose M Such. 2019. More than smart speakers: security and privacy perceptions of smart home personal assistants. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS)*. 451–466.
- [17] M. Billur Akdeniz, Roger J. Calantone, and Clay M. Voorhees. 2013. Signaling quality: An examination of the effects of marketing- and nonmarketing-controlled signals on perceptions of automotive brand quality. *Journal of Product Innovation Management* 31, 4 (2013), 728–743. <https://doi.org/10.1111/jpim.12120>
- [18] Angeliki Aktypi, Jason RC Nurse, and Michael Goldsmith. 2017. Unwinding Ariadne's identity thread: Privacy risks with fitness trackers and online social networks. In *Proceedings of the 2017 on Multimedia Privacy and Security*. 1–11.
- [19] Commercial Alert. 2008. Comments Before the Federal Communications Commission in the Matter of Sponsorship Identification Rules and Embedded Advertising. *MB Docket* 08-90 (2008).
- [20] Amazon. 2024. *All-new Echo Show 8 (3rd Gen, 2023 release) | With Spatial Audio, Smart Home Hub, and Alexa | Charcoal Brand: Amazon*. <https://www.amazon.com/All-New-Echo-Show-8/dp/B0BLS3Y632/>
- [21] Moritz Becker, Andreas Kolbeck, Christian Matt, and Thomas Hess. 2017. Understanding the Continuous Use of Fitness Trackers: A Thematic Analysis. *21ST PACIFIC-ASIA CONFERENCE ON INFORMATION SYSTEMS (PACIS 2017)*.
- [22] Sruti Bhagavatula, Lujo Bauer, and Apu Kapadia. 2022. "Adulthood is trying each of the same six passwords that you use for everything": The Scarcity and Ambiguity of Security Advice on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.
- [23] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
- [24] Virginia Braun and Victoria Clarke. 2021. Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and psychotherapy research* 21, 1 (2021), 37–47.
- [25] Joseph Bugeja, Désirée Jönsson, and Andreas Jacobsson. 2018. An investigation of vulnerabilities in smart connected cameras. In *2018 IEEE international conference on pervasive computing and communications workshops (PerCom workshops)*. IEEE, 537–542.
- [26] Rita Marie Cain. 2011. Embedded advertising on television: Disclosure, deception, and free speech rights. *Journal of Public Policy & Marketing* 30, 2 (2011), 226–238.
- [27] Andrew Tzer-Yeu Chen, Morteza Biglari-Abhari, and Kevin I-Kai Wang. 2018. Context is King: Privacy Perceptions of Camera-based Surveillance. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 1–6. <https://>

- //doi.org/10.1109/AVSS.2018.8639448
- [28] Kevin Childs, Daniel Nolting, and Anupam Das. 2023. Heat Marks the Spot: De-Anonymizing Users' Geographical Data on the Strava Heatmap. (2023).
  - [29] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the experts: What should be on an IoT privacy and security label?. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 447–464.
  - [30] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. 2020. Ask the Experts: What Should Be on an IoT Privacy and Security Label?. In *2020 IEEE Symposium on Security and Privacy (SP)*. 447–464. <https://doi.org/10.1109/SP40000.2020.00043>
  - [31] Pardis Emami-Naeini, Janarth Dheenadhayalan, Yuvraj Agarwal, and Lorrie Faith Cranor. 2021. Which privacy and security attributes most impact consumers' risk perception and willingness to purchase IoT devices?. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE, 519–536.
  - [32] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
  - [33] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. 2019. Exploring How Privacy and Security Factor into IoT Device Purchase Behavior (*CHI '19*). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300764>
  - [34] Chinonso E Etumnu, Kenneth Foster, Nicole O Widmar, Jayson L Lusk, and David L Ortega. 2020. Does the distribution of ratings affect online grocery sales? Evidence from Amazon. *Agribusiness* 36, 4 (2020), 501–521.
  - [35] Hossein Fereidooni, Tommaso Frassetto, Markus Miettinen, Ahmad-Reza Sadeghi, and Mauro Conti. 2017. Fitness trackers: fit for health but unfit for security and privacy. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE, 19–24.
  - [36] Nathaniel Fruchter and Ilaria Llicardi. 2018. Consumer attitudes towards privacy and security in home assistants. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.
  - [37] Sandra Gabriele and Sonia Chiasson. 2020. Understanding fitness tracker users' security and privacy knowledge, attitudes and behaviours. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
  - [38] Vaibhav Garg. 2021. A Lemon by Any Other Label. *ICISSP* (2021), 558–565.
  - [39] Christine Geeng and Franziska Roesner. 2019. Who's in control? Interactions in multi-user smart homes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
  - [40] Julie M Haney, Yasemin Acar, and Susanne Furman. 2021. "It's the Company, the Government, You and I": User Perceptions of Responsibility for Smart Home Privacy and Security.. In *USENIX Security Symposium*. 411–428.
  - [41] Julie M Haney, Susanne M Furman, and Yasemin Acar. 2020. Smart home security and privacy mitigations: Consumer perceptions, practices, and challenges. In *International Conference on Human-Computer Interaction*. Springer, 393–411.
  - [42] Hamza Harkous, Sai Teja Peddinti, Rishabh Khandelwal, Animesh Srivastava, and Nina Taft. 2022. Hark: A Deep Learning System for Navigating Privacy Feedback at Scale. In *2022 IEEE Symposium on Security and Privacy (SP)*.
  - [43] Gary Horcher. 2018. *Woman says her Amazon device recorded private conversation, sent it out to random contact*. <https://www.kiro7.com/news/local/woman-says-her-amazon-device-recorded-private-conversation-sent-it-out-to-random-contact/755507974/>
  - [44] Huafeng Jin and Shuo Wang. 2018. Voice-based determination of physical and emotional characteristics of users. US Patent 10,096,319.
  - [45] Alexandra Klausner. 2023. *I heard a strange man talk to my daughter in her room — our baby monitor was hacked*. <https://nypost.com/2023/10/23/lifestyle/parents-say-baby-monitor-was-hacked-someone-talked-to-child/>
  - [46] Kerri-Ann L Kuhn, Margee Hume, and Anita Love. 2010. Examining the covert nature of product placement: Implications for public policy. *Journal of Promotion Management* 16, 1-2 (2010), 59–79.
  - [47] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
  - [48] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–31.
  - [49] Jingjie Li, Kaiwen Sun, Brittany Skye Huff, Anna Marie Bierley, Younghyun Kim, Florian Schaub, and Kassem Fawaz. 2023. "It's up to the Consumer to be Smart": Understanding the Security and Privacy Attitudes of Smart Home Users on Reddit. In *IEEE Symposium on Security and Privacy (SP)(SP)*. IEEE Computer Society Los Alamitos, CA, 380–396.
  - [50] Chantal Lidynia, Philipp Brauner, and Martina Ziefle. 2018. A Step in the Right Direction – Understanding Privacy Concerns and Perceived Sensitivity of Fitness Trackers. In *Advances in Human Factors in Wearable Technologies and Game Design*, Tareq Ahram and Christianne Falcão (Eds.). Springer International Publishing, 42–53.
  - [51] Zhijie Lin, Ying Zhang, and Yong Tan. 2019. An empirical study of free product sampling and rating bias. *Information Systems Research* 30, 1 (2019), 260–275.
  - [52] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies* 2019, 4 (2019).
  - [53] Lydia Manikonda, Aditya Deotale, and Subbarao Kambhampati. 2018. What's up with privacy? User preferences and privacy concerns in intelligent personal assistants. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. 229–235.

- [54] Philipp Morgner, Christoph Mai, Nicole Koschate-Fischer, Felix Freiling, and Zinaida Benenson. 2020. Security update labels: establishing economic incentives for security patching of IoT consumer products. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 429–446.
- [55] Vivian Genaro Motti and Kelly Caine. 2015. Users' privacy concerns about wearables. In *International Conference on Financial Cryptography and Data Security*. Springer, 231–244.
- [56] George Orwell. 1950. 1984.
- [57] Joseph Phelps, Glen Nowak, and Elizabeth Ferrell. 2000. Privacy concerns and consumer willingness to provide personal information. *Journal of public policy & marketing* 19, 1 (2000), 27–41.
- [58] Zablon Pingo and Bhuvana Narayan. 2018. Users' responses to privacy issues with the connected information ecologies created by fitness trackers. In *International Conference on Asian Digital Libraries*. Springer, 240–255.
- [59] Elissa M Redmiles. 2019. "Should I Worry?" A Cross-Cultural Examination of Account Security Incident Response. In *2019 IEEE Symposium on Security and Privacy (SP)*. 920–934.
- [60] Alvin E. Roth. 2016. *Who gets what—and why: The New Economics of matchmaking and Market Design*. Mariner Books/Houghton Mifflin Harcourt.
- [61] Eva-Maria Schomakers, Chantal Lidynia, and Martina Zieffle. 2019. Listen to My Heart? How Privacy Concerns Shape Users' Acceptance of e-Health Technologies. In *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. 306–311. <https://doi.org/10.1109/WiMOB.2019.8923448>
- [62] Sajjad Shahid. 2022. *15 Top Ecommerce Websites That Are Massively Popular Among Shoppers in 2022*. <https://www.cloudways.com/blog/top-ecommerce-websites/>
- [63] Sergey Shekhan and Artem Hartutyunyan. 2013. Watching the watchers: hacking wireless IP security cameras. *Shape Security and Qualys Inc* (2013).
- [64] Michael Spence. 1973. Job market signaling. *The Quarterly Journal of Economics* 87, 3 (1973), 355. <https://doi.org/10.2307/1882010>
- [65] Adam Taylor. 2013. *Woman says her Amazon device recorded private conversation, sent it out to random contact*. <https://www.businessinsider.com/foscam-baby-monitor-hack-2013-8>
- [66] Christian Thorun, Max Vetter, Lucia A Reisch, and Anne K Zimmer. 2017. Indicators of consumer protection and empowerment in the digital world: Results and recommendations of a feasibility study. (2017).
- [67] Dirk van der Linden, Matthew Edwards, Irit Hadar, and Anna Zamansky. 2020. Pets without PETs: on pet owners' under-estimation of privacy concerns in pet wearables. *Proc. Priv. Enhancing Technol.* 2020, 1 (2020), 143–164.
- [68] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kevin Huguenin, and Mauro Cherubini. 2022. Are Those Steps Worth Your Privacy? Fitness-Tracker Users' Perceptions of Privacy and Utility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 181, 41 pages. <https://doi.org/10.1145/3494960>
- [69] Swaathi Vetrivel, Veerle van Harten, Carlos H Gañán, Michel van Eeten, and Simon Parkin. 2023. Examining Consumer Reviews to Understand Security and Privacy Issues in the Market of Smart Home Devices. (2023).
- [70] Jessica Vitak, Yuting Liao, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2018. Privacy Attitudes and Data Valuation Among Fitness Tracker Users. In *Transforming Digital Worlds*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, 229–239.
- [71] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End user security and privacy concerns with smart homes. In *Symposium on Usable Privacy and Security (SOUPS)*, Vol. 220.
- [72] Eric Zeng, Shrirang Mare, and Franziska Roesner. 2017. End User Security and Privacy Concerns with Smart Homes. In *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association, Santa Clara, CA, 65–80. <https://www.usenix.org/conference/soups2017/technical-sessions/presentation/zeng>
- [73] Shikun Zhang, Yuanyuan Feng, Lujo Bauer, Lorrie Faith Cranor, Anupam Das, and Norman Sadeh. 2021. "Did you know this camera tracks your mood?": Understanding Privacy Expectations and Preferences in the Age of Video Analytics. *Proceedings on Privacy Enhancing Technologies* 2021, 2 (2021), 282–304.
- [74] Shaohu Zhang, Zhouyu Li, and Anupam Das. 2023. VoicePM: A Robust Privacy Measurement on Voice Anonymity. In *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '23)*. 215–226.
- [75] Shikun Zhang and Norman Sadeh. 2023. Do privacy labels answer users' privacy questions. In *Network and Distributed System Security Symposium*.
- [76] Feng Zhu and Xiaoquan Zhang. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of marketing* 74, 2 (2010), 133–148.
- [77] Michael Zimmer, Priya Kumar, Jessica Vitak, Yuting Liao, and Katie Chamberlain Kritikos. 2020. 'There's nothing really they can do with this information': unpacking how users manage privacy boundaries for personal fitness information. *Information, Communication & Society* 23, 7 (2020), 1020–1037.

## A Details of the codebook for the enhanced keyword selection process

### A.1 Dataset

$\mathcal{D}'_1$ , a stratified subsample of  $\mathcal{D}_1$  which is representative of the reviews containing keywords ‘security’ and ‘privacy’ stratified across keywords, device types, and geographic regions.

### A.2 Variables

Two variables, *title* and *content*, had been selected. The variable *title* is the heading of a review. The variable *content* is the body of the review. Both are textual data.

### A.3 Coding scheme and objective

Values of the text variables *title* and *content* will be read with the aim of choosing candidate keywords to be included in the *enhanced-keyword-set*.

### A.4 Illustrated example

See Figure 5

### A.5 Missing data

We did not observe any reviews with missing data because the data preprocessing step already filtered out any reviews with empty *content*. However, in some cases, we observed missing *title* in a product review.

### A.6 Software and tools

Spreadsheets have been used for bookkeeping for datasets, codebooks, and compilation of the *enhanced-keyword-set*.

### A.7 Ethical considerations

The information about the consumer who posted the reviews had been anonymized and then circulated across coders to read. The reviews had been quoted as is. All typos were kept in their original forms. Longer reviews were preceded and followed by ellipses to show the relevant parts.

## B Details of codebook for labeling reviews containing S&P concern

### B.1 Dataset

$\mathcal{D}_2$ , the dataset which is representative of the reviews containing keywords from the *enhanced-keyword-set*.

### B.2 Variables

The same as Appendix A.2.

### B.3 Coding scheme and objective

Values of the text variables *title* and *content* will be read with the aim to label each review with a ‘1’ if the review has any security & privacy concerns and with ‘0’ otherwise.

### B.4 Illustrated Example

See Figure 6.

## B.5 Missing data

The same as Appendix A.5.

## B.6 Software and tools

Spreadsheets have been used for bookkeeping datasets, codebooks, and labeling each review.

## B.7 Ethical considerations

The same as Appendix A.7.

# C Details of codebook for the thematic analysis

## C.1 Dataset

The filtered rows of  $\mathcal{D}_2$  with label = 1, i.e., the reviews from  $\mathcal{D}_2$  that had been labeled to have security and privacy concerns.

## C.2 Variable definition

The same as Appendix A.2.

## C.3 Coding scheme and objective

Values of the text variables *title* and *content* will be read with the aim of finding out parts (called *excerpts*) with security & privacy concerns. Next, each *excerpt* will be coded into fine-grained subthemes related to security and privacy. Finally, similar fine-grained subthemes will be grouped together under a coarse-grained main theme.

## C.4 Illustrated example

See Figure 7

## C.5 Missing data

The same as Appendix A.5.

## C.6 Software and tools

Dedoose, a tool used for coding and thematic analysis, has been used for the coding process.

## C.7 Ethical considerations

The same as Appendix A.7.

# D Themes and subthemes found in the thematic coding process

## D.1 Fitness Tracker

- Issues with app(s):
  - Affiliation with third-party app(s)
  - Alexa
  - Being hacked
  - Existence of hack bots
  - Invasiveness
  - Lack of MFA
  - Permissions



- Requires an online account
- Requires to join a social network
- Running apps on the phone
- Scam(s)
- Concerns about hacking
- User error of product
- Company affiliation
  - Data shared
  - A tech company acquiring Fitbit
  - Requiring an account from a tech company
- Tracking
  - Allow tracking
  - Location services
  - Monitoring always
  - Always tracking
  - Concerns about location tracking
  - Uses gps
- Data security
  - Collect data
  - Data breach
  - Data share
- Payment security
  - Contactless payment feature
  - Credit card number/details
  - No payment system of their own
  - Not willing to pay with the device
- Wifi security
  - Requires wifi connection
  - Wifi password
- Invasive Customer support
  - Intrusiveness

## D.2 Smart Home Speaker

- Intrusion or Invasion
  - Always listening
  - General comment(s) on its intrusiveness
  - General comment(s) on its invasiveness
  - Getting targetted ads based on eavesdropping
  - Intrusive app(s)
  - Intrusive or invasive notification(s)
  - Listening (to conversations)
  - Making unauthorized connection(s) to other devic(es)
  - Records conversation(s)
  - Transcribe activit(ies)
  - Ads based on conversations

- Annoying ads
- Concerns of spying
- Concerns of spyware
- Eavesdropping
- Invasion of privacy
- Invasiveness
- Listens while off
- Monitoring/recording
- Records conversations
- Personal skepticism and opinions
  - Rhetorical question(s)
  - Sarcastically stating Alexa's privacy violating nature
  - Self-righteous or being indifferent about privacy
  - Skepticism on privacy or security aspect
  - Stating the user is the product
  - Taking privacy violation for granted
  - Concerns of privacy undermined by utility
  - Concerns with parent company
  - General privacy concerns
  - General security concerns
  - Skepticism / lack of trust
- Lack of trust in privacy control
  - Muting or turning off microphone
  - Putting it in a more public or safe space and not in private place like bedroom
  - Snooping nature of the mic
  - Turn off (unplug) for privacy
  - Using the camera shutter or Turning camera off for privacy in Echo Show
  - Concerns with privacy policy
- Data security
  - Account details for payments
  - Data collection (or retention)
  - General comments on data security
  - Lack of control over data or information
  - On data encryption
  - On requiring, sharing, or accessing personal and sensitive information
  - Ability to delete recording history
  - Concerns about transcripts
  - Data concerns
  - Remembers conversations
  - Selling user's data
- Technical and technological issues
  - Activates even when not called
  - Activating all the features to make the device stop spying
  - Delete record(s) for privacy
  - Drop in feature
  - Making subscription(s) without user's knowledge

- Password security
- Privacy feature(s) or setting(s) of the app(s)
- Privacy over voice commands
- Security updates
- Sidewalk feature
- Tech savvy comments
- Using the microphone
- Wifi security
- Concerns about privacy settings
- Concerns with devices communicating together
- Concerns with integrating into the home
- False activation
- Hearing other people on the device
- Privacy concerns while sharing prime
- Privacy concerns with drop in feature
- problems with Wi-Fi
- Security concerns related to Wi-Fi connection
- Self-activated
- Weird noises
- Surveillance by the government or some other regime(s)
  - Breach of federal law
  - Government or other government related agencies listening or spying
  - Referral to the novel 1984
  - Referring to a Surveilling Regime by Big Brother
  - Regulations government can impose against Amazon
  - Share information with government
  - Privacy concerns with government
- Concerns or vignettes about hackers
  - Vignette of being hacked
  - Worried about hackers or being hacked
  - Hacking concerns

### D.3 Surveillance and Security Camera System

- Data Security
  - Data collection
  - General comments on data security and data traffic
  - On Cloud Service or not being to work with a local storage
  - On sharing sensitive or personal information
  - Requires phone number to use the app
  - Requires sharing login details of one account among multiple users
  - Concerns with too much personal information
  - Concerns with uploading data to the cloud
- Issues with App(s) or Device
  - Account security
  - Authentication or password security

- Company require remote access to the device
- Historic or live video viewing protocol
- Lack of encryption
- Live view to another app only possible if mic is enabled at the camera
- Misleading notification by the app
- Permission(s) and consent(s)
- Privacy notice or policy
- Privacy related functionalities
- Receiving spams
- Records when not allowed
- Requires to be used on Google Chrome
- Requiring Internet to work
- Router or wifi or network security
- Shutter of the camera
- Third-party app(s) or feature(s)
- Camera always on
- Concerns with account sharing
- Concerns with password requirements
- Concerns with privacy settings
- Forcing users to create an account
- Malfunctioning product
- Required Wi-Fi connection
- Concerns or vignettes about hackers
  - Concerns about being hacked
  - Vignette of being hacked or someone taking control over the camera
  - Hacking concerns
- Personal skepticism and opinions
  - Ones with tech savvy jargon(s)
  - Sarcasm(s)
  - Skepticisms about the privacy and security aspect(s) of the product
  - Concerns with privacy policy
  - General privacy concerns
  - General security concerns
  - High levels of technical expertise
- Country affiliation
  - Data flow to a specific country
  - Lack of trust in the software or product from a specific country
- Lack of Trust in the device
  - Cover cameras for privacy
  - Turn off mic for privacy
  - Turn off or unplug for privacy
  - Concerns of spying
  - Invasiveness/intrusiveness
- Surveillance by a Regime
  - Referring to 1984
  - Referring to the regime as Big Brother

### E Proportion (in percentage) of signals by Keyword

This section contains three tables, one for each category of devices chosen. Each table has the same set of columns. A value in the column keyword is a keyword that was part of the *enhanced-keyword-set* for that category. A value in the column ‘Non-concern’ denotes in how many non-concern reviews that keyword was found. A value in the column ‘Concern’ denotes in how many concern reviews that keyword was found. The ‘Total’ column holds the values that are the sum of the previous two column values. The column proportion is the ratio expressed in percentage of the number of concern reviews in which a word was found to the total number of reviews in which the keyword was found. This percentage could be considered a proportion of signals for that keyword.

**Table 14.** Proportion (in percentage) of signals by keywords chosen for FT.

Keyword	Non-concern	Concern	Total	Proportion (%)
privacy	7	10	17	58.82
requires location	9	7	16	43.75
hacked	3	7	10	70
location services	22	7	29	24.14
personal data	38	6	44	13.64
intrusive	11	5	16	31.25
invasive	0	4	4	100
data security	4	4	8	50
card information	9	4	13	30.77
hacker	0	3	3	100
hackers	0	3	3	100
data privacy	5	3	8	37.5
invade privacy	0	2	2	100
privacy invading	0	2	2	100
privacy concerns	0	2	2	100
digital security	0	2	2	100
invade	1	2	3	66.67
invading	1	2	3	66.67
encrypted	0	1	1	100
goodbye privacy	0	1	1	100
invasive of privacy	0	1	1	100
privacy-violating	0	1	1	100
creepily	0	1	1	100
security concern	2	1	3	33.33
dubious	11	1	12	8.33

**Table 15.** Proportion (in percentage) of signals by keywords chosen for SHS.

Keyword	Non-concern	Concern	Total	Proportion (%)
privacy	278	351	629	55.8
recording	311	197	508	38.78
spy	52	156	208	75
spying	52	156	208	75
privacy concerns	36	86	122	70.49
intrusive	132	72	204	35.29
intrusion	133	72	205	35.12
privacy issues	36	63	99	63.64
privacy settings	87	63	150	42
invasion	23	57	80	71.25
invasive	23	57	80	71.25
personal privacy	17	34	51	66.67
personal information	104	31	135	22.96
invaded	10	23	33	69.7
invades	10	23	33	69.7
security concern	16	23	39	58.97
invaded privacy	3	21	24	87.5
collecting data	8	20	28	71.43
securing data	4	12	16	75
violation	9	11	20	55
safety	171	10	181	5.52
hackers	2	8	10	80
listened 24/7	14	8	22	36.36
cautious	17	7	24	29.17
security risk	0	6	6	100
threats	3	5	8	62.5
beware of privacy	2	4	6	66.67
illegal	4	4	8	50
network security concern	1	3	4	75
invades personal privacy	0	2	2	100
masquerading	0	2	2	100
wpa	1	2	3	66.67
wpa2	2	2	4	50
trojan	2	1	3	33.33

**Table 16.** Proportion (in percentage) of signals by keywords chosen for SSCS.

Keyword	Non-concern	Concern	Total	Proportion (%)
privacy	127	74	201	36.82
phone number	95	67	162	41.36
password	204	42	246	17.07
hacked	21	41	62	66.13
hacking	21	41	62	66.13
personal information	15	21	36	58.33
privacy issue	16	20	36	55.56
government	18	12	30	40
privacy rights	7	10	17	58.82
hackers	6	9	15	60
privacy mode	34	9	43	20.93
privacy reasons	4	7	11	63.64
violated	4	6	10	60
privacy policy	2	5	7	71.43
sell data	3	5	8	62.5
encrypts	6	5	11	45.45
intrusive	17	5	22	22.73
extort	19	3	22	13.64
permission	3	2	5	40
spied	4	2	6	33.33
moral code	0	1	1	100
infiltrate	1	1	2	50
strange noises	2	1	3	33.33
warrant	9	1	10	10
snoop	16	1	17	5.88



F Example mappings from excerpts to concerns that covers confidentiality, integrity, authentication, secrecy, anonymity, solitude

**Table 17.** Example mappings from excerpts to concerns that cover confidentiality, integrity, authentication, secrecy, anonymity, solitude.

Excerpt from review	Excerpt	Mapped to
FT-257- $\mathcal{D}_2$	Like for the pay thing, there's one more way to have a credit card number out-there, so when FitBit is hacked, there's another cyber thief to get such information	Confidentiality
SSCS-732- $\mathcal{D}_2$	...However, it is completely baffling that there can only be one user account for each system. You can share access with someone else, but that means giving them your username and password. It also means that everyone who has access to this account can delete videos or alter settings...	Integrity
FT-380- $\mathcal{D}_2$	Found out after purchasing that it needs to connect to wifi for all the features to work. (It also can not accept special characters in the wifi password. So you either have to live with only bluetooth connectivity and no music, or dumb down the security on your wifi.)	Authentication
SHS-1871- $\mathcal{D}_2$	Also, there is a setting where you can erase any of your words that were recorded in wake mode. But my fam isn't discussing the secret to perpetual motion, so we really don't worry about it. Any of us reading these reviews and are worried about the "government gathering secrets" or "letting the government in" can re-evaluate their use of computers, cell phones, data, and wifi.	Secrecy
SSCS-1192- $\mathcal{D}_2$	My research on this camera has revealed that all your network information such as IP address, gateway IP, SSID and password (for your wifi security) would be sent to and stored in a server somewhere in China. All videos captured by the camera would be sent as RTSP to that server (they call it "cloud"). And if you do not want to use that feature (cloud service), then the viewing and monitoring features would become useless (not functional).	Anonymity
SHS-1820- $\mathcal{D}_2$	"Now that might be fine for you, but that's not fine for me. I don't want my bedroom Dot to be connected to any other Alexa devices in the house. Imagine my surprise when I went into the phone app and found that anyone in the house could just "drop in." "Turn that off Amazon! That should be a very clear setting on initial config, with a warning that says "WARNING: DO NOT TURN THIS ON IF YOU DON'T WANT OTHERS IN THE HOUSE SNOOPING ON YOU."	Solitude or isolation

# G Visual Illustration of steps taken in the keyword selection, review labeling, and thematic coding processes.

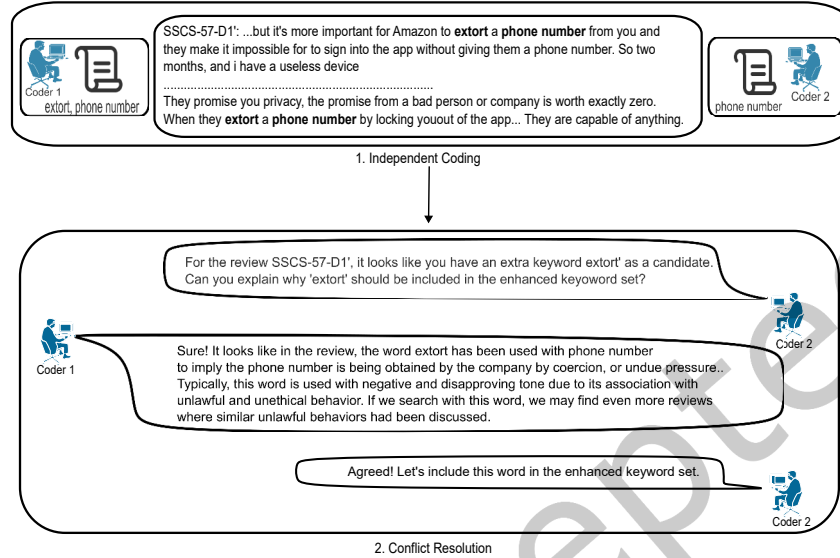


Fig. 5. The coding (Step 1) and the Conflict resolution (Step 2) in the keyword selection process.

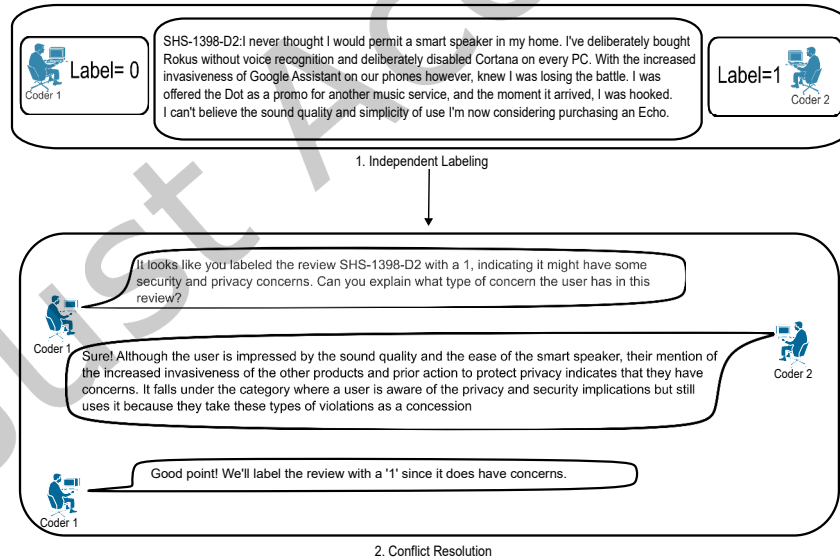


Fig. 6. The coding (Step 1) and the Conflict resolution (Step 2) in the review labeling process.

Received 23 May 2024; revised 19 July 2024; accepted 13 July 2024

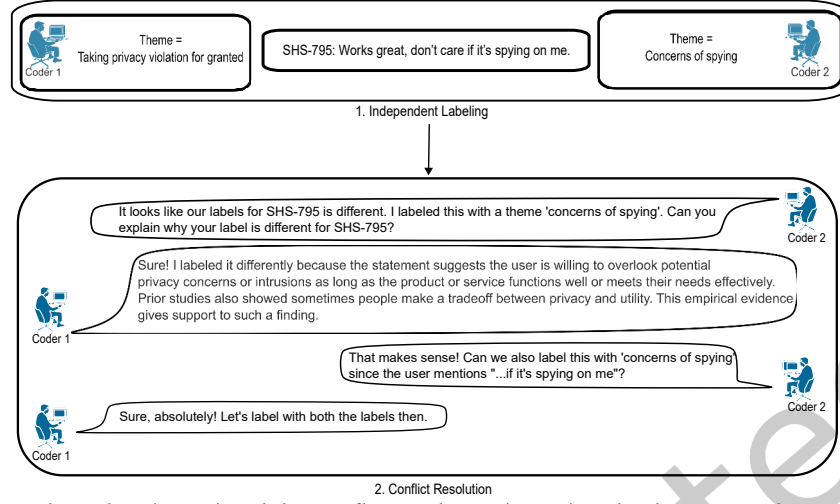


Fig. 7. The coding (Step 1) and the Conflict resolution (Step 2) in the thematic analysis process.

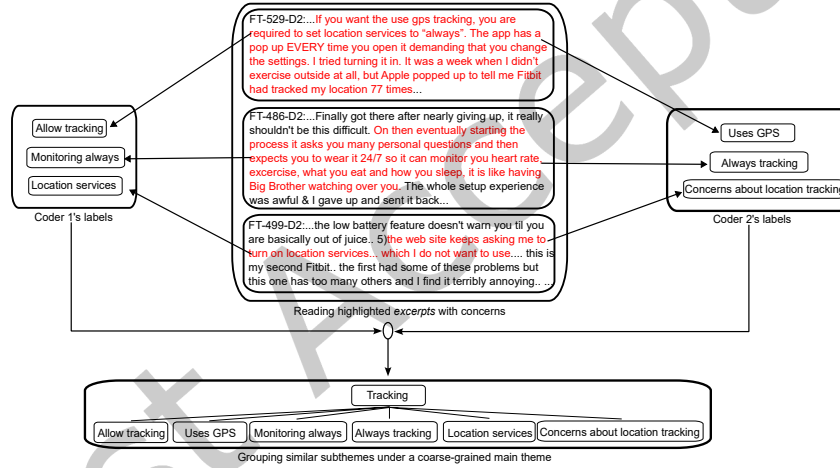


Fig. 8. The step of grouping similar fine-grained themes into a coarse-grained theme.