Protecting User Privacy in the World of Internet of Things

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Computer Science)

by

Tu Le

August 2023

Abstract

With the rapid development of the Internet of Things (IoT), there are many interacting devices and applications. They usually need to collect data about users and the surrounding environment to operate. With so many data collection activities going on daily, one of the most critical challenges is ensuring user privacy while facilitating devices' operations. In this dissertation, we aim to better understand the privacy risks and how to effectively protect user privacy in IoT environments. We first investigated the smart home context, which is one of the most popular IoT systems. One important feature of IoT for smart homes is voice-controlled devices/applications, which provide convenience but also introduce privacy issues. Unlike mobile apps, voice-controlled apps (or voice apps) do not have binary files or source code, posing a challenge to traditional analysis approaches. To address this challenge, we build a system to automatically interact with voice apps to identify suspicious behaviors and investigate privacy risks, such as inappropriate content or personal data collection. Although service providers like Amazon Alexa employed a vetting process for publishing voice apps, we find risky child-directed apps were still published on the market. To understand the root causes of such risks, we conduct a rigorous systematic analysis of voice apps' dynamic behaviors that bypassed standard vetting schemes and uncover limitations of the vetting process. We then propose a run-time behavior monitoring approach to address those privacy issues. In addition to risks from voice apps, there are also privacy concerns with voice-controlled devices and their data collection, as these devices record users' voice interactions. To study the impacts of this data collection and how to help users take control over it, we build a usable tool to effectively help users monitor their voice interactions recorded by the devices in real-time. We conduct user studies to understand users' perceptions of this data collection and how they prefer privacy notifications in the real world. It is important to also investigate public IoT systems and see if users' privacy perceptions and preferences vary. Thus, we further extend our research to the smart commercial building context. We conduct a user study with occupants who report working in smart commercial buildings regarding awareness of data collection, privacy notification preferences, and the potential factors for notification preferences, showing the key differences from the smart home context. Overall, our research helps to understand the risks of IoT technologies and informs key designs for smart environments to protect user privacy.

Approval Sheet

This dissertation is submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

Tu Le

Tu Le

This dissertation has been read and approved by the Examining Committee:

Yuan Tian

Yuan Tian, Advisor

Seongkook Heo

Seongkook Heo, Committee Chair

Yixin Sun

Yixin Sun, Committee Member

Sara Riggs

Sara Riggs, Committee Member

Danny Y. Huang

Danny Y. Huang, Committee Member

Accepted for the School of Engineering and Applied Science:

Jennifer L. West

Jennifer L. West, Dean, School of Engineering and Applied Science

August 2023

To my family and everyone who has been an influence.

Acknowledgments

I am fortunate to have Prof. Yuan Tian as my PhD advisor. I want to thank Yuan for the opportunity to be in her research lab, for her mentorship, and for her support in my work. She taught me a lot during my PhD study and guided me through the difficult times. I appreciate her hard work and last-minute help, which motivated me a lot. She steered my research in the right direction and helped me improve my skills. She also introduced me to potential collaborations for my projects.

I want to thank Prof. Seongkook Heo, Prof. Yixin Sun, Prof. Sara Riggs, and Prof. Danny Y. Huang for serving as my dissertation committee members and giving constructive feedback on my work. Danny, in particular, collaborated with me on most of my projects throughout my PhD journey and taught me a lot. I would like to also thank him for his understanding and support when I went through the tough times in my life.

I am also thankful to my fellow graduate students, including Jianfeng Chi, Fnu Suya, Faysal Hossain Shezan, Tamjid Al Rahat, William Shand, Kunlin Cai, and my other collaborators.

My PhD research would not have been possible without the great support from the Department of Computer Science at the University of Virginia and the Department of Electrical and Computer Engineering at the University of California, Los Angeles.

Finally, I am grateful to my family for their unconditional support in my life.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Internet of Things (IoT) market is predicted to reach 19 billion devices in 2025 and grow to 29 billion devices by 2030 [1]. With the rapid development of the Internet of Things (IoT), there are many interacting devices and applications, such as autonomous vehicles, smart homes, and smart cities. These devices and applications consist of sensing, computing, actuation, and communication. Applications can interact with different entities, including users, devices, other applications, and the environment. For example, in smart home environments, we have IoT devices such as security cameras, smart locks, smart speakers, smart meters, and smoke detectors, and they are all connected to the Internet. There are also larger-scale IoT systems in smart commercial buildings, industrial facilities, or smart cities where hundreds or thousands of devices are working together. Such systems require a lot of data collection about users and the surrounding environment to operate. The interaction with both the physical world and the Internet through smart devices has increased the attack surface dramatically and led to a lot of vulnerabilities. For example, voice control features suffer from adversarial attacks on speech recognition systems. With so many IoT devices and applications collecting data about users, one of the most critical challenges is ensuring user privacy while allowing data collection in an IoT world.

In this dissertation, our goal is to study how to protect user privacy in an IoT world. To achieve this goal, we conduct studies to better understand the privacy issues and user expectations in IoT environments. We also introduce new tools to protect user privacy. Specifically, we investigate voice-controlled apps' privacy issues, different modalities of privacy notification for voice data collection, and smart building occupants' perceptions of IoT data collection and notification preferences. Our studies show that **we can protect user privacy in the world of IoT by bridging the gap between software behaviors and users' privacy preferences**.

We first look into the smart home context, which is one of the most popular IoT systems. One important feature of

IoT for smart homes is the ability to interact with the devices and control them via voice commands. Many devices include a microphone and a speaker to receive voice commands from users and answer queries through voice. Though this is a convenient and interactive way to improve user experience, it also raises security and privacy issues, especially from voice apps developed for these voice-controlled devices. Prior work mostly examined security issues of app hijacking and limitations of voice recognition systems. Little work was done on how to automatically analyze the voice apps at scale, what the existing problems are with the current app market, and how they particularly affect consumers. There are some critical challenges to systematically evaluating voice apps (hereinafter also referred to as skills). Although many prior studies have performed measurements and compliance analysis for mobile apps, little large-scale automated analysis has been done for the voice app domain. Unlike mobile apps, traditional analysis approaches do not apply well to these voice apps because such apps' backend is hosted on cloud servers, and these cloud servers can be from third parties. Furthermore, this backend is dynamic and malicious behaviors might not always occur in every interaction with the apps. Therefore, automating the testing and analysis to catch such malicious behaviors is challenging as it requires generating valid inputs and dynamic follow-up inputs to reveal possible hidden behaviors following certain templates, which could not be achieved using existing NLP techniques for chatbots. To address these challenges, we build a system to automatically interact with voice apps and analyze the resulting interactions. Our system enables tracking suspicious behaviors of voice apps and investigating privacy risks, such as inappropriate content or personal information collection through voice interactions, in voice apps' logic.

Second, although service providers like Amazon Alexa employed a vetting process to check the policy compliance of the voice apps, our above system still identifies violating apps getting published on the market, including child-directed apps. Several previous studies also found non-compliant apps and malicious behaviors [2, 3, 4]. However, it is unclear how the standard vetting process actually works and how such violating apps could still get published. To further investigate the root causes, we propose a rigorous systematic analysis of voice apps' behaviors, focusing on dynamic behavior change. We reverse engineer the Alexa vetting process and uncover the limitations that lead to policy-violating voice apps remaining published. To address these privacy issues, we present a run-time app behavior monitoring approach to protect user privacy when users interact with voice apps. Our proposed solution aims to ensure the validity and consistency of the skill's behaviors at run-time without introducing much overhead.

Third, in addition to the risks from voice apps, there are privacy concerns with the voice-controlled devices' data collection, as these devices record users' interactions. To understand users' awareness of such data collection and identify their concerns, we conduct a survey to understand voice assistant users' awareness of their voice data recording and their expectations for monitoring their voice data. There is currently a lack of tools to assist users with effectively controlling such voice interaction data collection. Based on the insights from our survey, we build a fast, easy-to-use browser extension to help users continuously monitor the recording of voice interactions done by voice-controlled devices (or its software known as voice personal assistants) and notify users of any suspicious recordings. Notifying users of the risks is an important task that needs to be designed carefully to ensure user experience. We then conduct a semi-structured interview study to explore users' preferences for different notification modalities and how users make decisions against the risks in the real world.

Finally, it is important to also investigate large-scale public IoT environments to identify key differences in users' privacy perceptions and preferences as compared to private smart homes. Therefore, we further extend our research to investigate the smart commercial building context. We envision that the future infrastructure will include many smart commercial buildings with many connected devices, stakeholders, and pervasive data collection. Despite the numerous potential benefits of making environments "smarter", the transition to a large-scale IoT world with many potential privacy concerns also introduces great challenges to human-centered smart system designs. Continuous data collection can expose more data than anticipated by the users, and the collected data can be shared with third parties [5]. One particular privacy issue in these environments relates to users' perceptions and preferences of data collection and practices. Disclosing data collection practices and notifying users of any issues is enforced by regulations. However, how to appropriately handle privacy notifications in smart buildings remains an unanswered question. For example, the power dynamics in smart commercial buildings (e.g., employers vs. employees) may influence how occupants perceive privacy as compared to private homes. In addition, smart building data collection is multi-modal, pervasive, and large-scale. Traditional privacy notification strategies might not fit well. It is important to explore building occupants' notification preferences to avoid overwhelming them or causing misunderstanding. Previous studies mostly focused on general privacy preferences in smart homes. However, key differences between smart homes and public settings such as smart commercial buildings are underexplored. To understand users' perceptions of IoT data collection and notification preferences in large-scale IoT environments, we conduct a user study with occupants who report working in smart commercial buildings regarding: 1) awareness and perception of data collection, 2) privacy notification preferences, and 3) potential factors for the privacy notification preferences.

Chapter 2

Background

This chapter presents related research in the area, including privacy risks of the voice-controlled devices and users' privacy preferences.

2.1 Privacy Risks of Voice-controlled Applications

Voice personal assistants (VPA) remain among the most popular IoT devices and have been a hot topic of security/privacy studies in prior work [6, 7, 8]. Given the ability to allow controlling by voice commands, VPAs were found to be vulnerable to security attacks and user privacy issues, mostly from speech recognition weaknesses [3, 9, 10, 11, 12, 13, 14]. For example, an adversary could impersonate the VPA or hijack third-party apps' sessions to eavesdrop on users. Several defense mechanisms were then proposed to address such issues [15, 16].

Previous work also looked into privacy practices and policy enforcement for voice apps (or widely known as skills). Edu et al. [4] showed that many published skills had bad privacy practices and did not implement permission requests properly. Many skills did not provide a valid privacy policy [17]. Other work investigated the Alexa skill vetting process and revealed some limitations that led to the fact that some violating skills being published [2, 18, 19].

VPAs (e.g., Alexa and Google Home) are basically a black box. Based on previous findings [20, 21, 22], users have an incomplete understanding of how VPAs work. Users are also concerned about their privacy and controls of their data [23, 24]. Other work further explored the privacy norms for VPAs in smart homes and applied contextual integrity to visualize information flows [25]. Pins et al. [26] designed an interactive approach to visualize interactions with VPAs and their data flows.

2.2 Privacy Preferences and Decision Making

There has been considerable literature investigating the privacy preferences and factors that affect users' privacy decision-making in IoT scenarios. Naeini et al. [27] showed that participants were more comfortable with data collection in public rather than private settings and were more likely to share data for uses that they find beneficial (e.g., find public restrooms). The collection of biometric data is considered less comfortable than environmental data, and the participants wanted to be notified about the data practices of such information being collected. In contrast, via in situ studies, previous work found users' privacy concerns or misunderstanding of the facial recognition technology [28, 29] and fitness tracker [30].

Other work looked into the influence of friends and experts on privacy decisions [31]. These studies showed that participants were more influenced when their friends denied data collection than when their friends allowed data collection. In contrast, the participants were more influenced when experts allowed collection than when experts denied data collection. However, after being exposed to a set of scenarios in which friends and experts allowed or denied data collection, the participants were less likely to be influenced in subsequent scenarios. Barbosa et al. [32] presented machine learning models to predict personalized privacy preferences in smart homes and identify factors that could change such preferences.

Several frameworks were proposed to help users to enforce privacy protections on IoT devices. Apthorpe et al. [33] presented a framework to discover the privacy norms in the smart home context. IoTWatch [34] allows users to specify their privacy preferences at install time and ensure IoT apps' behaviors match the selections. Kratos [35] provides smart home users with access control settings that consider multiple users and devices in a shared space.

2.3 Privacy Notifications for IoT

Privacy notifications are a type of privacy notice that informs people or users about the data being collected and using practices of a system, product, or service. They are often provided by entities responsible for the disclosed data practices (e.g., data collection, sharing, and processing), as increasing requirements by privacy regulations around the world (e.g., GDPR [36], CCPA [37]). Although privacy policies are the most common type of privacy notice, they are lengthy and difficult to read [38, 39]. Instead, researchers and practitioners proposed more effective ways to notify people about data privacy practices, such as concise privacy notices [40] and privacy nutrition labels [41]. Schaub et al. also outlined a design space for more effective privacy notices [42].

User-facing privacy notifications, in addition to or in lieu of privacy policies, are very common in the digital world. For example, websites have increasingly adopted GDPR-compliant cookie banners, which automatically pop up when websites detect new users. These banners usually contain a concise privacy notification describing how the website uses cookies to track user data and how users can disable some of them [43]. Another common example is the app permission management framework on smartphones. Both iOS and Android platforms send users just-in-time notifications when apps are trying to access certain sensitive permission on smartphones, along with the choice to allow or deny. The notifications in both examples are delivered to users through primary channels, which is the same platform or device a user interacts with. Huang et al. [44] presented a tool that examines network traffic in a smart home and informs the user of vulnerabilities or tracking services.

However, in IoT embedded smart buildings, providing people with effective privacy notices are extremely challenging. First, since smart buildings have countless IoT devices and sensors collecting data (e.g., energy sensors, lighting, temperature, air quality, etc.), it is possible that the number of notifications can overwhelm building occupants and visitors. This may lead to privacy fatigue [45] if users receive too many irrelevant privacy notices. Second, IoT devices and sensors in smart buildings lack traditional user interfaces (e.g., screens), so it is difficult to deliver privacy notifications through the most intuitive primary channels (i.e., IoT devices and sensors themselves). This means privacy notices need to rely on secondary or public channels (e.g., a website, physical signs), causing an additional barrier for residents or visitors to receive them. Researchers have recently developed a location-based mobile app, IoT Assistant, capable of notifying people about nearby IoT data privacy in public places [46, 47]. However, little research has examined what types of IoT privacy notices people would like to receive and how to receive them.

Designing more comprehensive and user-friendly privacy dashboards can be an effective way to give users more control over their security and privacy when using voice personal assistants (VPAs). The importance and effectiveness of privacy dashboards were highlighted by Irion et al. [48] as one of the most feasible methods of enhancing control for users and maintaining consistency with rising standards of privacy. Creating a well-designed privacy dashboard is complex, however, requiring an understanding of user demands as well as general best design practices. Farke et al. [49] surveyed users of Google's My Activity dashboard to better understand user perceptions and reactions to a privacy dashboard. Raschke et al. [50] designed a mock-up dashboard to present a possible implementation for generalized sensitive data. However, Feth et al. [51] emphasized that privacy dashboards should be tailored to the specific domain and technology. Thus, a privacy dashboard for voice personal assistants should be designed and tested independently. Sharma et al. [52] surveyed users of Google Voice Assistant in particular to explore the specific needs and expectations of VPA technology and controls while also designing an algorithm to classify sensitive VPA interactions. However, these prior studies have yet to design a working privacy dashboard with enhanced features and test it on real users.

Chapter 3

Systematical Voice-controlled Apps' Content and Risk Analysis

3.1 Overview & Goal

In this chapter, we looked into privacy issues of voice apps (hereafter referred to as "skills") developed for voice-controlled devices. Particularly, we proposed techniques to systematically analyze the content of voice apps and identify the potential risks. Systematically evaluating voice apps at scale to identify those that contain risky content or confounding utterances was a challenging task. Unlike mobile apps, voice apps are hosted on their service provider's servers. This makes empirical analysis of VPA skills challenging, as neither the executable files nor source code of VPA skills is available to researchers. Our goal was to automate the skill behaviors collection and analysis of skills to identify risky content without requiring manual voice interactions. We specifically investigated the skills made for children (referred to as "kid skills"). We also explored parents' attitudes and awareness of the risks.

3.2 Introduction

The rapid development of Internet of things (IoT) technology has aligned with growing popularity of voice personal assistant (VPA) services, such as Amazon Alexa and Google Home. In addition to the first-party features provided by these products, VPA service providers have also developed platforms that allow third-party developers to build and publish their own voice apps—hereafter referred to as "skills".

Risks to Children from VPAs. Researchers have found that 91% of children between ages 4 and 11 in the U.S. have access to VPAs, 26% of children are exposed to a VPA between 2 and 4 hours a week, and 20% talk to VPA devices more than 5 hours a week [53]. The lack of robust authentication on commercial VPAs makes it challenging to regulate children's use of VPA skills [54], especially as anyone in the physical vicinity of a VPA can interact with the device. Additionally, parental control modes provided by VPAs (e.g., Amazon FreeTime and Google Family App) often place a burden on parents during setup and receive complaints from parents due to their limitations [55, 56, 57].

Legal and industry efforts have tried to protect children using VPAs; however, their effectiveness is unclear. In the U.S., the 1998 Children's Online Privacy Protection Act (COPPA) places information collected online from children under the age of 13 under parental control [58], but widespread COPPA violations have been shown in the mobile application market [59], and compliance in the VPA space is far from guaranteed. Other work has studied parents' concerns about their children's Internet use [60], including reports that more than 80% of U.S. parents are concerned about their children having access to sexual content, making friends with strangers, and exposing personal information [61].

In this study, we consider two types of threats posed by VPAs to children: (1) risky skills and (2) confounding utterances. We define "risky" skills as skills that contain inappropriate content for children or ask for personal information through voice interactions. An example is the "My Burns" skill in Amazon's Kids category that says "You're so ugly you'd scare the crap out of the toilet." Another example is a skill named "Shape Game" that asks for age information: "Awesome! Before we start however; I'm curious...how old are you?" We define "confounding utterances" as voice commands shared among two or more skills that could cause a user to unintentionally invoke a different skill than intended.

Anyone can create a developer account and publish their skills to the Alexa Skills store for free. Additionally, the Alexa Skills Kit developer console [62] makes it easy to develop, validate, and publish skills even with limited experience. Therefore, risky skills and confounding utterances could be the result of intentionally malicious developers or benign/inexperienced developers unaware of potential risks.

Challenges to Automated Skill Analysis. It is challenging to systematically evaluate VPA skills to identify those that contain risky content or confounding utterances. Substantial prior research has focused on COPPA compliance [63, 64] and large-scale measurements [65] in the mobile application domain. For example, Reyes et al. [63, 64] showed that many mobile apps on the Google Play Store and Apple App Store appeared to violate COPPA and infringe on privacy rights. Razaghpanah et al. [65] similarly identified and characterized organizations associated with mobile advertising and user tracking.

However, little large-scale automated evaluation has been performed in the VPA domain. Unlike mobile apps that users download and install on their smartphones, VPA skills are essentially web apps hosted on VPA service providers' cloud servers. Therefore, VPA skills do not provide binary files or installation packages for download. This makes empirical analysis of VPA skills challenging, as neither the executable files nor source code of VPA skills are available to researchers. Most existing techniques and frameworks for automated analysis of mobile applications employ static or dynamic analysis of binary or source code and cannot be applied to VPA skills.

VPA skills' natural language processing modules and key function logic are hosted in the cloud as a black box. VPA skill voice interactions are built following a template defined by the third-party developer, which is also unavailable to researchers. To automatically detect risky content, we need to generate testing inputs that trigger this content through sequential interactions. A further challenge is that risky content does not always occur during a user's first interaction with a skill; human users often need to have back-and-forth conversations with skills to discover risky content. Automating this process requires developing a tool that can generate valid voice inputs and dynamic follow-up responses that will cause the skill to reveal risky content. This is different from existing chatbot techniques [66], as the goal is not to generate inputs that sound natural to a human. Instead automated skill analysis requires generating inputs that explore the space of skill behaviors as thoroughly as possible.

Research Questions. The risks and challenges discussed above mean that protecting children in the era of VPAs raises several pressing questions:

- **RQ0.** Can we automate the analysis of VPA skills to identify risky content for children without requiring manual voice interactions?
- **RQ1.** Are VPA skills specifically targeted to children that claim to follow additional content requirements – hereafter referred to as "kid skills" – actually safe for child users?
- **RQ2.** What are parents' attitudes and awareness of the risks posed by VPAs to children?
- **RQ3.** How likely is it for children to be exposed to risky skills through "confounding utterances"—voice commands shared by multiple skills that could cause a child to accidentally invoke or interact with a different skill than intended.

In this study, we address the challenges to analyzing skills at scale. We design, implement, and perform a systematic automated analysis of the Amazon Alexa VPA skill ecosystem. We then conduct a user study to validate our results and further answer these research questions.

Contributions. We make the following contributions:

Automated System for Skill Analysis (RQ0). We present a natural-language-based system, "SkillBot," that automatically interacts with Alexa skills and collects their contents at scale (Section 3.5). SkillBot generates valid skill inputs, analyzes skill responses, and systematically generates follow-up inputs. SkillBot can be run longitudinally to identify new conversations and new conversation branches in previously analyzed skills. We made our project available to the public to help facilitate future research.

Identification of Risks to Children (RQ1). We take a systematic approach to analyze VPA skills based on automated interactions. We analyze 3,434 Alexa skills specifically targeted toward children in order to measure the prevalence of kid skills that contain risky content, including inappropriate language and personal information collection (Section 3.6). Through multiple rounds of interactions, we identify 28 kid skills with risky contents and maintain a growing dataset of 31,966 non-overlapping skill behaviors.

User Study of Parents' Awareness and Experiences (RQ2). We aim to better understand the real world contexts of children's interactions with VPAs and verify our SkillBot results by seeing whether parents also viewed identified skills as risky. We conduct a user study of 232 U.S. Alexa users who have children under 13 years old (Section 3.7). A majority of surveyed parents express concern about the content of the risky kids skills identified by SkillBot. Many also express disbelief that these skills are actually available for Alexa VPAs. This lack of risk awareness is compounded by findings that many parents do not use VPA parental controls and allow their children to use VPA versions that do not have parental controls enabled by default.

Confounding Utterances (RQ3). We identify confounding utterances as a novel threat to VPA users (Section 3.8). Confounding utterances are voice commands shared by more than one skill that may be present on a VPA device. When a user interacts with a VPA via a confounding utterance, it might trigger a reaction from any of these skills. If a kid skill shares a confounding utterance with a skill inappropriate for kids, a child user might inadvertently interact with the inappropriate skill. For example, a child could use a confounding utterance to invoke a kid skill X only to have non-kid skill Y triggered instead. As many VPAs do not offer visual cues of what skill is actually invoked, the user may not realize that skill Y is running instead of skill X. Our analysis reveals 4,487 confounding utterances shared between two or more skills and highlights those that place child users at risk of invoking a non-kid skill instead of a kid skill.

3.3 Related Work

We discuss related work in this area and compare this prior research to our project. The related work can be categorized into three main streams as follows.

3.3.1 Assessing Personal Assistants and Voice Apps

Previous work looked into the problem of speech recognition misinterpretations made by voice personal assistants [13, 14, 67], showing that an adversary could impersonate the voice assistant system or other skills to eavesdrop on users. Edu et al. [4] showed that many developers implemented permission requests with bad privacy practices. Other work reversed engineered the Alexa skill vetting process and revealed limitations that allowed policy-violating skills to be published [2, 68]. Lentzsch et al. [2] and Liao et al. [17] also found that many skills did not provide a valid privacy policy. Unlike these previous papers, we investigate the hidden logic inside skills to identify risky content. Recently, Guo et al. [69] investigated skills asking users' for private information by triggering the skills starting with the three sample utterances provided by the skills themselves.

Our work focuses on children's safety when using voice apps, including a broader set of risks than explored by [69], including inappropriate content and asking for personal information. Our SkillBot also explores the possible conversations that users can have with Alexa skills in much greater depth and breadth via our chatbot module. This is critical for identifying risky skill behaviors, especially in kid skills that have to follow stricter policies to get published. For example, out of the 28 risky kid skills (8 expletive and 20 sensitive) identified in our study, 3 required the use of custom utterances generated by Skillbot. In addition, 2 risky skills were only identified at a depth of 11 in the conversation tree, 1 skill at depth 5, 4 skills at depth 4, 6 at depth 3, 8 at depth 2, and 7 at depth 1. Skillbot's ability to conduct such extended back-and-forth interactions with skills separates it from prior work in the area. We also ran a user study to understand parents' experiences, awareness, concerns about risky content for children, and use of parental control. In addition, we described and analyzed confounding utterances—a novel threat which exposes children to risky contents.

3.3.2 Perceptions of Personal Assistants

Major et al. [20] found that Alexa users often confused third-party skills with built-in Alexa features, and they did not know what features the native Alexa system supports. Abdi et al. [21] found that users have different perceptions of how data is processed by personal assistants and that users have security/privacy concerns about what the assistants learn. They further explored the privacy norms for personal assistants in smart homes and visualized the acceptability of information flows based on contextual integrity [25]. Pins et al. [26] designed an interactive approach to visualize conversations with voice assistants and help users with data literacy. Meng et al. [24] investigated the perceptions of ownership for personal assistants in multi-user smart homes and showed that most users felt like an owner of the shared device although they did not have the same rights or controls. Lau et al. [23] found non-users did not trust the smart speaker companies while users expressed few concerns but did not completely understand the risks. They also highlighted that users often trade privacy for convenience and the current privacy controls are not effective. Different from these prior works, our user study focuses on the impacts of risky skills on children. We find that children often use types of Echo devices other than the Kids Edition in their households and that parents have concerns about the risky skills we identified. Additionally, many parents do not think such risky skills are available to children although those skills are actually published for children.

3.3.3 Children and the Internet

There is considerable literature examining children's online protection and COPPA compliance in domains other than VPA. Automated analyses of thousands of mobile applications have found widespread COPPA violations [59, 70] ranging from illegal collection of persistent identifiers to non-compliant privacy policies. Investigations of children-directed websites have also found covert tracking techniques designed to avoid COPPA requirements [71], among other non-compliant behaviors [72, 73]. Our work sheds light on the new problems in VPA. More recently, researchers have raised concerns about children's privacy with respect to Internet-connected "smart" toys. Several studies have conducted detailed analyses of specific toys, often noting multiple implementation decisions and security vulnerabilities that place children's data at risk and violate COPPA [74, 75, 76, 77]. Other studies have provided frameworks for smart toy protections [78, 79], and recommendations for smart toy manufacturers. As VPAs straddle the boundary between IoT products, mobile application platforms, and search engines, our work contributes to the evolving landscape of risks posed to children by modern online services, joining this related literature to demonstrate the breadth of challenges remaining.

Importantly, technical research into children and the Internet has been motivated and supported by qualitative and quantitative user studies investigating how children and parents understand connected technologies [80] and make privacy decisions. These studies show that parents are becoming more concerned about smart toy privacy [81, 82], but that children have difficulty conceptualizing certain types of privacy risks [83, 84]. Given that some parents actively compromise their children's online privacy [85] or help their children avoid COPPA protections [86], we should not rely on parents to keep children safe online. Instead, risky practices, such as the skills we identify, must be addressed through a combination of academic, regulatory, and industry action.

3.4 Threat Model

In this study, we consider two main types of threats: (1) risky skills (i.e., skills that contain inappropriate content or ask for users' personal information through voice interaction) and (2) confounding utterances (i.e., utterances that are shared among two or more different skills). This section explains these two threats.

Risky Skills. We investigate skills containing risky content that may harm children. We define "risky" skills as containing either or both of two kinds of content: (1) inappropriate content for children or (2) requests for personal information through voice interaction. An example of inappropriate content is the "My Burns" skill in Amazon's Kids category that says "You're so ugly you'd scare the crap out of the toilet." Another example is a kid skill named "Shape Game" that asks for age information: "Awesome! Before we start however; I'm curious...how old are you?" Note that these risks may be subjective. Thus, we evaluate our findings via a user study to understand the impacts of these risky skills. These threats may come from an adversary who intentionally develops malicious skills or a benign/inexperienced developer who is not aware of the risks.

Confounding Utterances. We identify a new risk which we call "confounding utterances". We define confounding utterances as voice commands that are shared among two or more different skills. Effectively, a confounding utterance could trigger an unexpected skill for the user.

Confounding utterances are different from previous research on voice squatting attacks, which exploited the speech recognition misinterpretations made by voice personal assistants [13, 14, 67, 87]. They showed that voice command misinterpretation problems due to spoken errors could yield unwanted skill interactions, and an adversary can route users to malicious Alexa skills by giving the malicious skills invocation names that are pronounced similarly to legitimate ones.

In contrast, we consider a new risk that exists even if there is no such voice command misinterpretation: Alexa may invoke a skill that the user did not intend if multiple skills are configured to respond to identical (confounding) utterances. We want to find out, given a confounding utterance that is shared between multiple skills, which skill Alexa prioritizes to enable/invoke. This matters because users have no control over which skill is actually opened when Alexa recognizes an intentional or unintentional voice command containing the confounding utterance (e.g., if Alexa is triggered by background conversations). The risk of confounding utterances is further exacerbated by Alexa's name-free interaction feature [88], which allows Alexa to invoke a skill even if a user does not explicitly state its invocation name. Furthermore, the lack of a download or installation process when a new skill is invoked makes it easy for these skills to bypass user awareness. In summary, a confounding utterance may cause a user to invoke a skill other than the one they intended. For instance, a child may have one skill in mind but accidentally invoke a different skill that responds to confounding utterances. An adversary could therefore exploit confounding utterances to get children to use malicious skills.

3.5 SkillBot - Automated Interaction with Skills

To study the impacts that risky skills might have on children, we propose SkillBot, which systematically interacts with skills to discover risky content and confounding utterances (**RQ0**). In this section, we first show how we design SkillBot for interacting with skills and collecting their responses thoroughly and at scale. We then evaluate SkillBot for its reliability, coverage, and performance.

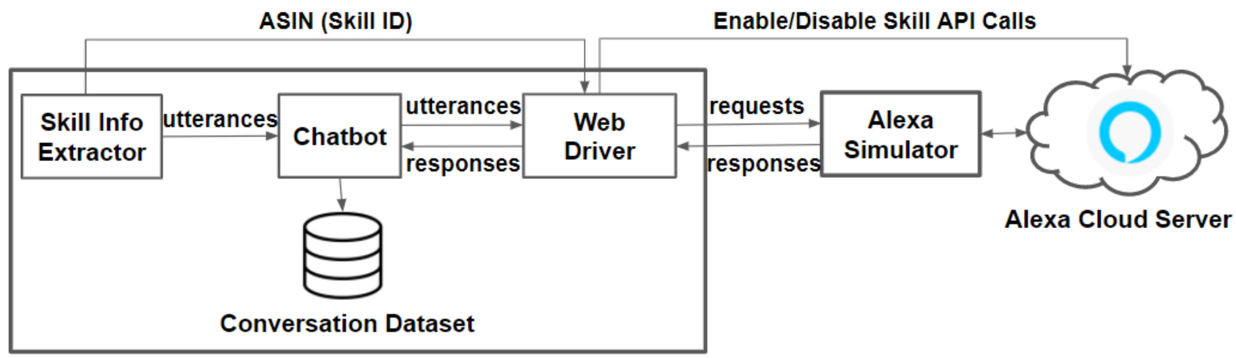


Figure 3.1: Automated skill interaction pipeline.

3.5.1 Automated Interaction System Design

Our goal is for SkillBot to interact effectively and efficiently with Alexa skills to uncover risky content for children in the skills' behaviors thoroughly and at scale.

Overview. Our system consists of four main components: (1) skill information extractor, (2) web driver, (3) chatbot, and (4) conversation dataset (Figure 3.1). The skill information extractor handles exploring, downloading, and parsing information about skills available in the Alexa Skills Store. The web driver handles connections to Alexa and requests from/to the skills. The chatbot discovers interactions with the skills and records the conversations into the conversation dataset.

Skill Information Extractor. Amazon provides an online repository of skills via the Alexa Skills Store [89]. Each skill is an individual product, which has its own product info page and an Amazon Standard Identification Number (ASIN) that can be used to search for the skill in Amazon's catalogue. The URL of a skill's info page can be constructed from its ASIN. Our skill information extractor includes a web scraper to systematically access the Alexa website and download the skills' info pages in HTML based on their ASINs. It then reads the HTML files and constructs a JSON dictionary structure using the BeautifulSoup library [90]. For each skill, we extract any information available on its info page, such as ASIN (skill ID), icon, sample utterances, invocation name, description, reviews, permission list, and category (e.g., kids, education, smart home, etc.).

Web Driver. We leverage Amazon's Alexa developer console [91] to allow programmatic interactions with skills using text inputs. Our web driver module uses the Selenium framework, which is a popular web browser automation framework for testing web applications, to automate sending requests to Alexa and interacting with the skill info page to check the status of the skill (i.e., enabled, disabled, not available). We also implement a module that handles skill enabling/disabling requests. This module uses private APIs derived from inspecting XMLHttpRequests within network activities of Alexa webpages.

Chatbot. Our NLP-based chatbot module interacts with the skills and explores as much of the skills' content as possible. The module includes several techniques to explore sample utterances suggested by the skill developers, create additional utterances based on the skill's info, classify utterances, detect questions in responses, and generate follow-up utterances. In the following paragraphs, we provide details of how our chatbot module explores and classifies utterances, detects questions, and generates follow-up utterances.

Exploring and Classifying Utterances: Amazon allows developers to list up to three sample utterances on their skill's information page. Our system first extracts these sample utterances. Some developers also put additional instructions into their skill's description. Therefore, our system further processes the skill's description to generate more utterances. In particular, we consider sentences that start with an invocation word (i.e., "Alexa, ...") to be utterances. We also notice that phrases inside quotes can be utterances. An example is "You can say 'give me a fun fact' to ask the skill for a fun fact." Once a list of collected utterances is constructed, our system classifies these utterances into opening and in-skill utterances. Opening utterances are used to invoke/open a skill. These often include the skill's name and start with opening words such as "open," "launch," and "start" [92]. In-skill utterances are used within the skill's session (when the skill is already invoked). Some examples include "tell me a joke," "help," or "more info." Figure 3.2 shows the workflow of how our chatbot module prepares potential utterances to test a skill.

Detecting Questions in Skill Responses: To extend the conversation, our system first classifies responses collected from the skill into three main categories: yes/no questions, WH questions, and non-question statements. For this classification task, we employ spaCy [93] and StanfordCoreNLP [94, 95], which are popular tools for NLP tasks. In particular, we first tokenize the skill's response into sentences and each sentence into words. We then annotate each sentence using part-of-speech (POS) tagging, including both TreeBank POS tags [96] and Universal POS tags [97]. The POS tags allow us to identify the role of each word in the sentence (e.g., auxiliary, subject, or object). A yes/no question usually starts with an auxiliary verb, which follows the subject-auxiliary inversion formation rule. Yes/no

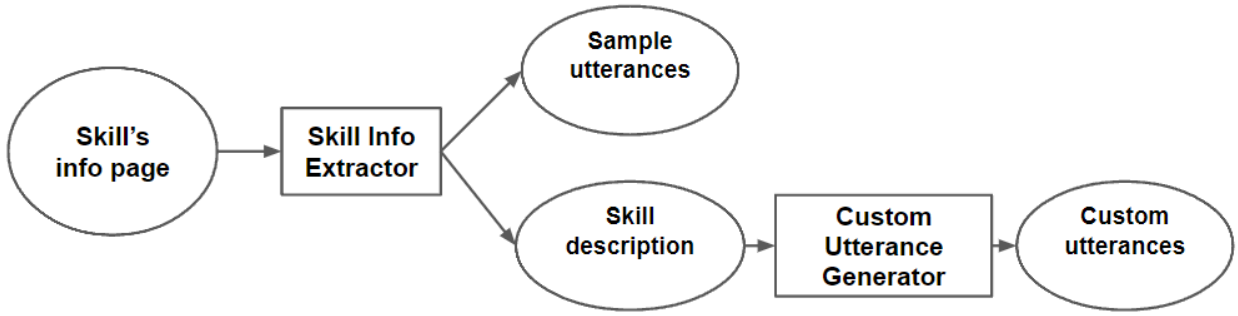


Figure 3.2: Workflow of exploring potential utterances to test a skill. Our chatbot module collects a variety of utterances, including sample utterances from the skill's info page and custom utterances constructed based on the skill description.

questions generally take the form of [auxiliary + subject + (main verb) + (object/adjective/adverb)?]. Some examples are "Is she nice?" "Do you play video games?" and "Do you swim today?" It is also possible to have the auxiliary verb as a negative contraction such as "Don't you know it?" or "Isn't she nice?" A WH question contains WH words such as "what," "why," or "how." These WH words can be identified in the sentence based on their POS tags: WDT, WP, WP$, and WRB. Regular WH questions usually take the form of [WH-word + auxiliary + subject + (main verb) + (object)?]. Some examples are "What is your name?" and "What did you say?" Furthermore, we consider pied-piping WH questions such as "To whom did you send it?" We exclude cases in which WH words are used in a non-question statement, such as "What you think is great," "That is what I did," or "What goes around comes around." Figure 3.3 shows our chatbot module's workflow for categorizing responses from skills. The chatbot then continues the conversations with the skills by generating follow-up utterances based on these categories.

Generating Follow-up Utterances: Given a skill response, our chatbot follows up based on the category of the response. (1) Yes/no questions. These questions ask for confirmation from the user, expecting either a "yes" or a "no" answer. The chatbot sends "yes" or "no" as a follow-up utterance to continue the conversation. (2) WH questions. The chatbot responds to WH questions based on the theme of the question. We use the classification method presented in [98] to categorize WH questions into one of six themes: abbreviation, entity, description, human, location, and numeric [99]. Abbreviation questions ask about a short form of an expression (e.g., "What is the abbreviation for California?"). Entity questions ask about objects that are not human (e.g., "What is color is the sky?"). Description questions ask about explanations of concepts (e.g., "What does a defibrillator do?"). Human questions ask about individuals or groups of people. Location questions ask about places, such as cities, countries, states, etc. Numeric questions ask about numerical values, such as count, weight, size, etc. Each question theme can also be divided into subthemes. Our chatbot contains a dictionary of answers to specific subthemes (e.g., "human:age": [1, 2, 3, ...], "location:states": [Oregon, Arizona, ...]) that can be used to continue the conversation with the skill. For questions

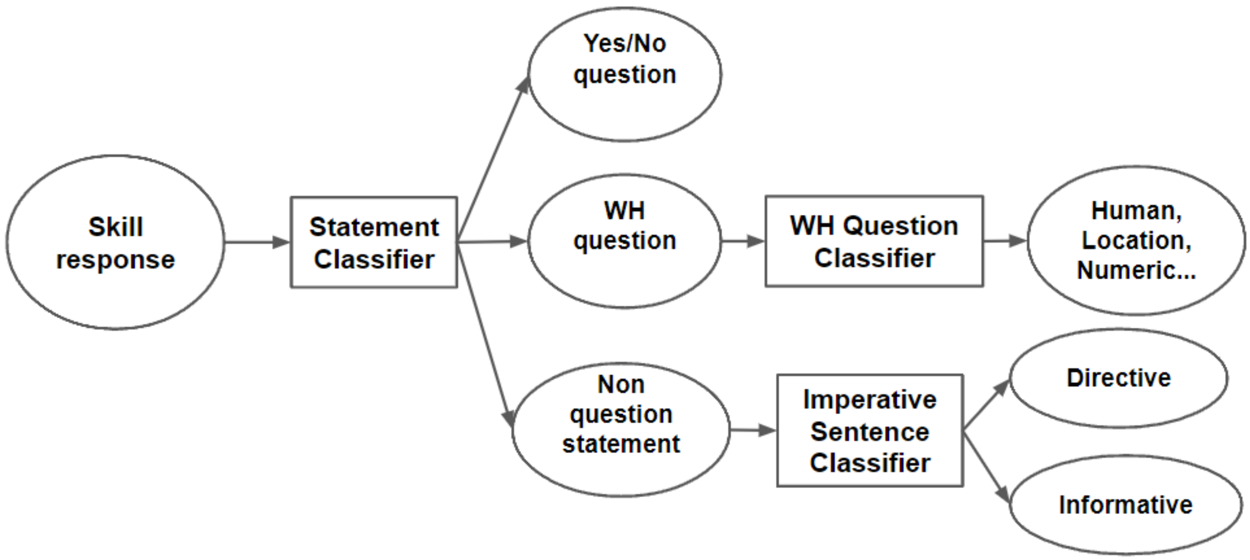


Figure 3.3: Our chatbot module's workflow for classifying a skill's responses in order to generate follow-up utterances.

in subthemes that are too general (including many in the description theme), the chatbot replies with "I don't know. Please tell me." to prompt for further responses from the skill. (3) Non-question statements. These include two types of statements: directive statements and informative statements. Some directive statements ask the user to provide an answer, which is similar to a WH question (e.g., "Please tell us your birthday"). For these cases, the chatbot parses the sentence to identify what being asked and responds as it would to a WH question. Other directive statements suggest words/phrases for the user to select to continue the conversation. Some examples include "Please say 'continue' to get a fun fact" and "Say '1' to get info about a book or '2' to get info about a movie." In these cases, the chatbot extracts the suggested words/phrases and uses them to continue the conversation. Informative statements provide users with some information, such as a joke, a fact, or daily news. These statements often do not indicate what the user can say to continue the conversation. The chatbot therefore sends an in-skill utterance "Tell me another one" or "Tell me more" as a follow-up to explore more content from the skill.

Conversation Dataset. Our conversation dataset is a set of JSON files, each of which represents a skill. The files contain a list of conversations with each skill collected by the chatbot module. Each conversation is stored as a list in which even indexes are the utterances sent by SkillBot and odd indexes are the corresponding responses from the skill.

3.5.2 Skill Conversation Data Collection

We ran SkillBot to collect conversation data from Alexa skills at scale. We explain SkillBot's detailed workflow for this process in the following paragraphs.

For each skill, SkillBot conducts multiple round of interactions to explore different *paths* within the *conversation tree*. Each node in this tree is a unique response from Alexa. There is an edge between nodes i and j if there exists an interaction in which Alexa says i, the user (i.e., SkillBot) says something, and then Alexa says j. We call the progression from i to j a *path* in the tree. Multiple paths of interactions can exist for a skill. For instance, node i could have two edges: one with j and another one with k. Effectively, two paths lead from i. In one path, the user says something after hearing i, and Alexa responds with j. In another path, the user says something else after hearing i, and Alexa responds with k.

To illustrate how we construct a conversation tree for a typical skill, we show a hypothetical example in Figure 3.4. First, the user would launch a skill by saying "Open Skill X" or "Launch Skill X." This initial utterance could be found in the "Sample Utterances" section of the skill's information page; alternatively, it could also be displayed in the "Additional Instructions" section on the skill's page. Following the example in Figure 3.4, let us assume that either "Open Skill X" or "Launch Skill X" triggers the same response from Alexa: "Welcome to Skill X. Say 'Continue'." This response is denoted by node 1 in Figure 3.4. The user would say "Continue" and trigger another response from Alexa (node 2): "Great. Would you like to do A?" The user could either respond with "Yes," which would trigger the response in node 3, or "No," which would trigger node 4.

SkillBot explores multiple paths of the conversation tree by interacting with a skill multiple times, picking a different response each time. Following the example in Figure 3.4, SkillBot could follow the path along nodes 1, 2, and 3 the first time it evaluates this skill. Once at node 3, the skill does not provide the user with the option to return to node 2, so SkillBot would have to start over to explore a different path. In the second run, SkillBot could follow a path along nodes 1, 2, 4, and 5. SkillBot responds with "No" after node 2 because it remembers answering "Yes" in the previous run. In the third run, SkillBot could follow nodes 1, 2, 4, and 6.

Each run of SkillBot terminates when exploring down a particular path is unlikely to trigger new responses from Alexa; in this case, SkillBot starts over with the same skill and explores a different path. We list four conditions where SkillBot would terminate a particular run: (1) Alexa's response is not new; in other words, SkillBot has seen the response repeatedly. SkillBot's goal is to maximize the interaction with unique skill responses in order to discover risky contents. (2) Alexa's response is empty. (3) Alexa's response is a dynamic audio clip (e.g., music or podcast) that does not rely on Alexa's automated voice. Due to limitations of the Alexa simulator, SkillBot is unable to extract and parse dynamic audio clips. Therefore, SkillBot terminates a path if it sees a dynamic audio clip because it does not know how

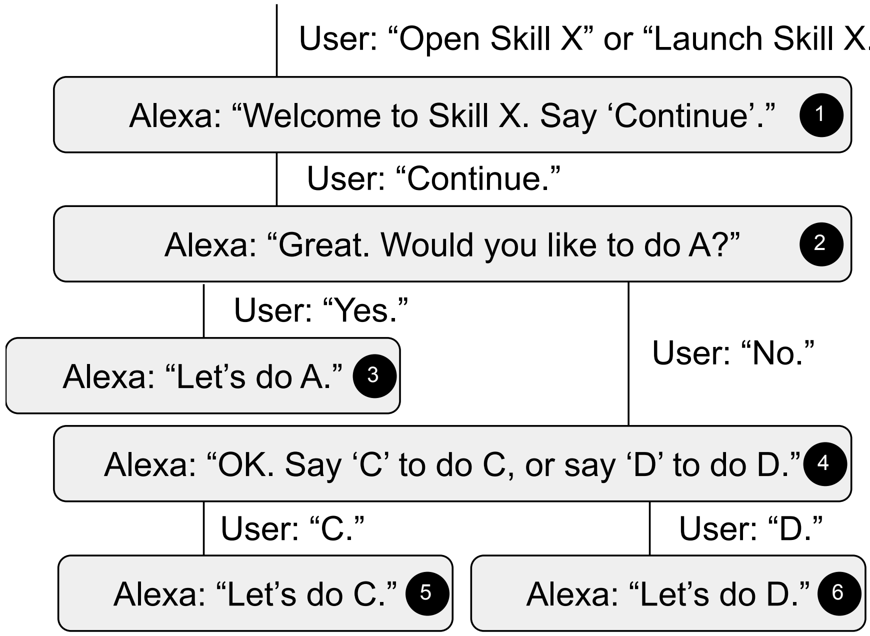


Figure 3.4: A conversation tree that represents how SkillBot interacts with a typical skill.

to react. (4) Alexa’s response is an error message, such as “The service is unavailable.” or “Sorry, I don’t understand.”

3.5.3 Evaluation

In this section, we present our validation to ensure that interacting with skills via SkillBot (Section 3.5) can represent a user’s interaction with skills via a physical Echo device.

Interaction Reliability We randomly selected 100 skills for validation. We used an Echo Dot device to manually interact with the skills and compared the responses against those collected by Skillbot. If the responses did not match, we further checked the skill invocation in the Alexa activity log to see if the same skill was invoked. We found that Skillbot and the Echo Dot have similar interactions across 99 of the 100 selected skills. Among these 99 skills, two skills responded with audio playbacks, which are not supported by the Alexa developer console [100] employed by SkillBot (Section 3.9). However, their invocations were shown in the activity log, which matched those invocations when using the Echo Dot. We cannot verify the one remaining skill as Alexa cannot recognize its sample utterances, potentially due to an issue with the skill’s web service.

Skill Response Classification. As described in Section 3.5.1, SkillBot extends the conversation with a skill by classifying the skill’s responses as yes/no questions, WH questions, and non-question statements. To evaluate the performance of this classification, we randomly sampled 300 unique skill responses from our conversation collection and manually labeled them as ground truth. This ground truth set included 52 yes/no questions, 50 WH questions, and 198 non-question statements. We then used SkillBot to label these responses and verified the labels against the ground truth. SkillBot predicted 56 yes/no questions, 50 WH questions, and 194 non-question statements, which is over 95% accuracy. The performance details for each class is shown in Table 3.1 (see Table A.1 in Appendix A.3 for the confusion matrix of our 3-class classifier).

	Accuracy	Precision	Recall	F1 Score
Yes/No	98%	0.91	0.98	0.94
WH question	98%	0.94	0.94	0.94
Non-question	96%	0.98	0.96	0.97

Table 3.1: Skill response classification performance

Coverage. We measure SkillBot's coverage by analyzing the collected conversation trees for every skill. Each skill can have multiple conversation trees representing different conversations. Our analysis includes four criteria: (1) the number of unique responses from Alexa, i.e., the number of nodes in a tree; (2) the maximum depth (or height) in a tree; (3) the maximum number of branches in a tree, i.e., how many options that SkillBot explored; and (4) the number of initial utterances, which counts the number of distinct ways to start interacting with Alexa. We show the results in Figure 3.5.

Per the second chart in Figure 3.5, SkillBot is able to reach a conversation tree depth of at least 10 on 2.7% of the skills. Such a depth allows SkillBot to trigger and explore a wide variety of Alexa's responses from which to discover risky contents. In fact, out of the 28 risky kid skills we identify (Section 3.6), 2 skills were identified at depth 11, 1 skill at depth 5, 4 skills at depth 4, 6 at depth 3, 8 at depth 2, and 7 at depth 1.

Per the fourth chart in Figure 3.5, SkillBot is able to initiate conversations with skills using more than 3 different utterances. Normally, a skill's information page lists *at most* three sample utterances. In addition to using these sample utterances, SkillBot also discovers and extracts utterances in the "Additional Instructions" section on the skill's page. As a result, SkillBot interacted with 20.3% of skills using more than 3 utterances. These extra initial utterances allow SkillBot to trigger more responses from Alexa. As we will explain in Section 3.6, 3 out of the 28 risky kid skills were discovered by SkillBot from additional utterances not listed among the sample interactions.

Time Performance. It took about 21 seconds on average to collect one conversation. SkillBot interacted with 4,507 skills and collected 39,322 conversations within 46 hours using five parallel processes on an Ubuntu 20.04 machine with an Intel Core i7-9700K CPU.

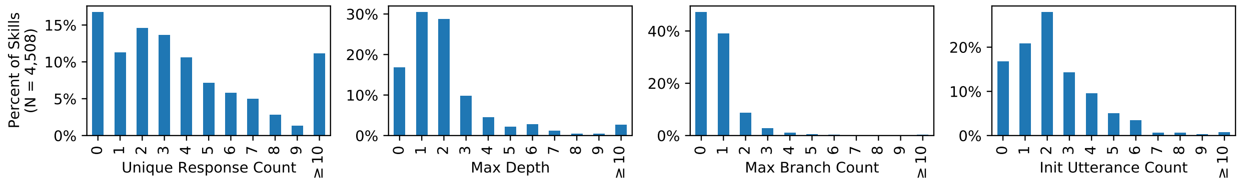


Figure 3.5: Coverage of SkillBot in terms of four criteria: number of unique responses from Alexa; maximum depth in a conversation tree; maximum number of branches for any node in a conversation tree; and number of initial utterances.

3.6 Kid Skill Analysis

To investigate the risks of skills specifically targeted to children (**RQ1**), we employed SkillBot to collect and analyze 31,966 conversations from a sample of 3,434 Alexa kid skills. In this section, we describe our dataset of kid skills and present our findings of risky kid skills.

3.6.1 Dataset

Our system first downloaded information about skills from their info pages on the Alexa U.S. skills store. We filtered out error pages (e.g., 404 not found) after three retries as well as non-English skills. As a result, we collected 43,740 Alexa skills from 23 different skill categories (e.g., business & finance, social, kids, etc.). Our system then parsed data about the skills, such as ASIN (i.e., skill's ID), icon, sample utterances, invocation name, description, reviews, permission list, and category, from the downloaded skill info pages.

For our analysis, we investigated all skills in Amazon's Kids category (3,439 kid skills). We ran SkillBot to interact with each skill and record the conversations. To speed up the task, we ran five processes of SkillBot simultaneously. Note that SkillBot can be run multiple times for each skill to cumulatively collect new conversations and new conversation branches for that skill. As a result, our sample had 31,966 conversations from 3,434 kid skills after removing five skills that resulted in errors or crashed Alexa.

3.6.2 Risky Kid Skill Findings

We performed content analysis on the conversations collected from 3,434 kid skills to identify risky kid skills that have inappropriate content or ask for personal information. Examples of risky skills identified in our analysis are attached in Appendix A.2.

Skills with Inappropriate Content for Children. Our goal was to analyze the skills' contents to identify risky skills that provide inappropriate content to children. To identify inappropriate content, we combined WebPurify and

Microsoft Azure's Content Moderator, which are two popular content moderation services providing inappropriate content filtering for websites and applications with a focus on child protection [101, 102]. We implemented a content moderation module for SkillBot in Python 3, leveraging the WebPurify API and the Azure Moderation API to flag skills that have inappropriate content for children.

The Skillbot content moderation module flagged 33 potentially risky skills with expletives in their contents. However, a human review process was necessary to verify this output, because whether or not a flagged skill's content is actually inappropriate for children depends on context. For example, some of the expletives (such as "facial" and "sex") are likely appropriate in some conversational contexts. For the human review process, four researchers on our team—who come from 3 countries (including the USA), all of whom are English speakers, and whose ages range from 22 to 35—independently reviewed each of the flagged skills and voted whether the skills' content is inappropriate for children. Skills that received three or four votes were considered to actually be inappropriate. Using this approach, we identified 8 kid skills with inappropriate content. Some examples include the "New Facts" skill that said "Here's your fact: A pig's orgasm lasts for 30 minutes" and the "My Burns" skill that said "You're so ugly you'd scare the crap out of the toilet." Out of these 8 kid skills, SkillBot identified the inappropriate content of one skill at conversation tree depth 11, one at depth 5, two at depth 4, one at depth 2, and three at depth 1.

We also sampled 100 other skills that were not flagged as having inappropriate content by SkillBot and manually checked them. We did not find inappropriate content in any of these skills, further validating SkillBot's performance.

Skills Collecting Personal Information. Our goal was to detect if kid skills asked users for personal information. To the best of our knowledge, available tools only focus on detecting personal information in the given text input, which is a different goal. For this analysis, we employed a keyword search to identify skill responses that asked for personal information. We constructed a list of personal information keywords based on the U.S. Department of Defense Privacy Office [103] and searched for these keywords in the skill responses. The keywords include name, age, address, phone number, social security number, passport number, driver's license number, taxpayer ID number, patient ID number, financial account number, credit card number, date of birth, and zipcode. However, a naive keyword search would not be sufficient, because in some cases the sentence containing any of those keywords might not actually ask for such information. We therefore combined keyword search and our SkillBot's question detection technique (presented in Section 3.5.1) to detect if a skill asked the user to provide personal information.

22 risky kid skills were flagged as asking users for personal information. To verify the result, we manually checked these 22 skills and 100 random skills that were not flagged. The manual verification found 2 false positives and 0 false negatives. Thus, 20 kid skills asked for personal information, such as name, age, and birthday. Some examples include

the "Ready Freddy" skill that kept asking kids to introduce themselves or the "Birthday Wisher" skill that asked for birthday info. Table A.2 in Appendix A.4 presents the confusion matrix for the evaluation.

Out of these 20 skills, SkillBot identified the queries for sensitive information in one skill at conversation tree depth 11, two skills at depth 4, six at depth 3, seven at depth 2, and four at depth 1. SkillBot also identified these queries in non-sample utterances for three of the skills (i.e., utterances listed in the "Additional Instructions" section of the skill's info page on Amazon.com rather than as sample interactions with the skill).

We further analyzed the permission requests of these 20 risky kid skills and found that none requested any permissions from the user.

3.7 Awareness & Opinions of Risky Kid Skills

To evaluate how the risky kid skills we identified actually impact child users (**RQ2**), we conducted a user study of 232 U.S. parents who use Amazon Alexa and have children under 13. Our goal was to qualitatively understand parents' expectations and attitudes about these risky skills, parents' awareness of parental control features, and how risky skills might affect children. Our study protocol was approved by our Institutional Review Board (IRB). The full text of our survey instrument is provided in Appendix A.1. In this section, we describe our recruitment strategy, survey design, response filtering, and results.

3.7.1 Recruitment

We recruited participants on Prolific,[1] a crowdsourcing website for online research. Participants were required to be adults 18 years or older who are fluent in English, live in the U.S. with their children under 13, and have at least one Amazon Echo device in their home. We combined Prolific's pre-screening filters and a screening survey to get this niche sample of participants for our main survey. Our screening survey consisted of two questions to determine (1) if the participant has children aged 1–13 and (2) if the participant has Amazon Echo device(s) in their household. 1,500 participants took our screening survey and 258 qualified for our main survey. The screening survey took less than 1 minute to complete and the main survey took an average of 6.5 minutes (5.2 minutes in the median case). Participants were compensated $0.10 for completing the screening survey and $2 for completing the main survey. To improve response quality, we limited both the screening and main surveys to Prolific participants with at least a 99% approval rate.

[1] https://www.prolific.co/

3.7.2 Screening Survey

The screening survey consisted of two multiple-choice questions: "Who lives in your household?" and "Which electronic devices do you have in your household?" This allowed us to identify participants with children aged 1–13 and Amazon Echo device(s) in their household who were eligible to take the main survey.

3.7.3 Main Survey

The main survey consisted of the following four sections.

Parents' Perceptions of VPA Skills. This section investigated parents' opinions of and experiences with risky skills. Participants were presented with two conversation samples collected by SkillBot from each of the following categories (six samples total). Conversation samples were randomly selected from each category for each participant and were presented in random order.

- *Expletive.* Conversation samples from 8 skills identified in our analysis that contain inappropriate language content for children.
- *Sensitive.* Conversation samples from 20 skills identified in our analysis that ask the user to provide personal information, such as name, age, and birthday.
- *Non-Risky.* Conversation samples from 100 skills that did not contain inappropriate content for children or ask for personal information.

The full list of skills in the Expletive and Sensitive categories are provided in Appendix A.2. Each participant was asked the following set of questions after viewing each conversation sample:

- Do you think the conversation is possible on Alexa?
- Do you think Alexa should allow this type of conversation?
- Do you think this particular skill or conversation is designed for families and kids?
- How comfortable are you if this conversation is between your children and Alexa?
- If you answered "Somewhat uncomfortable" or "Extremely uncomfortable" to the previous question, what skills or conversations have you experienced with your Alexa that made you similarly uncomfortable?

Amazon Echo Usage. We asked which device model(s) of Amazon Echo our participants have in their household (e.g., Echo Dot, Echo Dot Kids Edition, Echo Show). We also asked whether their kids used Amazon Echo at home.

Awareness of Parental Control Feature. We asked the participants if they think Amazon Echo supports parental control (yes/no/don't know). Participants who answered "yes" were further asked to identify the feature's name (free-text response) and if they used the feature (yes/no/don't know).

Demographic Information. At the end of the survey, we asked demographic questions about gender, age, and comfort level with computing technology. Our sample consisted of 128 male (55.2%), 103 female (44.4%), and 1 preferred not to answer (0.4%). The majority (79.7%) were between 25 and 44 years old. Most participants in our sample are technically savvy (68.5%). Table 3.2 shows detailed demographic information of our participants.

	Responses	Percentage
Gender		
Male	128	55%
Female	103	44%
Prefer not to answer	1	<1%
Age		
18 - 24	3	1%
25 - 34	61	26%
35 - 44	124	53%
45 - 54	40	17%
55 - 64	3	1%
65 and above	1	<1%
Comfort level with computing technology		
Ultra Nerd	19	9%
Technically Savvy	159	68%
Average User	54	23%
Luddite	0	0%

Table 3.2: Demographic information (gender, age, comfort level with computing technology) of the participants in our sample. The "Responses" column contains the number of participants who selected the corresponding choices. Our sample is nearly gender-balanced with most participants in the 25–44 age group. Most participants also self-reported to be technically savvy or average users.

3.7.4 Response Filtering

We received 237 responses for our main survey. We filtered out responses from participants who incorrectly answered either of two attention check questions ("What is the company that makes Alexa?" and "How many buttons are there on an Amazon Echo?"). We further reviewed the submissions and excluded participants who gave meaningless responses (e.g., straight-lining or entering only whitespaces into all free-text answer boxes). This resulted in 232 valid responses for analysis.

3.7.5 User Study Results

We find that most parents allow their children to use types of Amazon Echo devices other than the Kids Edition. Such types of Echo devices do not have parental control enabled by default. We also find that many parents do not know about the parental control feature. For those who know about the feature, only a few of them use it. Thus, many children potentially have access to risky skills. Our results further show that parents are not aware of the risky skills that are available in the Kids category on Amazon. When presented with examples of risky kid skills that have expletives and those that ask for personal information, parents express concerns, especially for skills with expletives. Some parents reported previous experiences with such risky skills.

Parents' Perceptions of Kid Skills. Table 3.3 shows the distribution of responses to the following questions across the Expletive, Sensitive, and Non-Risky skill sets:

- Do you think the conversation is possible on Alexa?
- Do you think Alexa should allow this type of conversation?
- Do you think this particular skill or conversation is designed for families and kids?

A majority of parents thought that the interactions with the Expletive skills were not possible and should not be allowed by Alexa. Only 45.9% of the parents thought these interactions were possible and only 41.6% thought such skills should be allowed. Furthermore, the majority of parents did not think that the Expletive skills were designed for families and kids.

The parents' responses with regard to the Expletive skills are significantly different from their responses to the Sensitive and Non-Risky skills on these questions. For each of these three questions, we conduct Chi-square tests on the pairs of responses across the skill sets: Non-Risky vs. Expletive, Non-Risky vs. Sensitive, and Expletive vs. Sensitive. The responses from the Expletive set are significantly different from responses from the other two sets for all three questions ($p < 0.05$). The responses to the "Alexa should allow" question are also significantly different for the Non-Risky set versus for the Sensitive set ($p < 0.05$). In contrast, the responses for the "Possible on Alexa" and "Designed for families and kids" questions display no significant difference between Sensitive and Non-Risky sets. This is alarming, as the Sensitive skills ask for personal information through the conversations with users, thereby bypassing Amazon's built-in permission control model for skills. As many skills are hosted by third parties, sensitive information about children could be leaked to someone other than Amazon.

Designed for Family and Children. Table 3.4 shows the distribution of responses for the question: "Do you think this particular skill or conversation is designed for families and kids?" with a breakdown across different types of skills

Response	Possible on Alexa			Alexa should allow			Designed for families and kids		
	Non-Risky	Expletive	Sensitive	Non-Risky	Expletive	Sensitive	Non-Risky	Expletive	Sensitive
Yes	78.0%	45.9%	71.6%	83.2%	41.6%	66.2%	68.1%	27.4%	55.8%
No	7.5%	30.2%	11.2%	8.8%	44.2%	16.6%	13.8%	57.1%	16.6%
Not sure	14.4%	23.9%	17.2%	8.0%	14.2%	17.2%	18.1%	15.5%	27.6%

Table 3.3: Distribution of responses for each of the three yes/no questions across three types of conversations.

Response	Expletive	Sensitive	Non-Risky
Yes	27.4%	55.8%	68.1%
No	57.1%	16.6%	13.8%
Not sure	15.5%	27.6%	18.1%

Table 3.4: Distribution of responses for the question: "Do you think this particular skill or conversation is designed for families and kids?"

(e.g., Non-Risky, Expletive, and Sensitive). Most parents (57.1%) felt that skills with expletives were not designed for families and kids. In addition, 15.5% were not sure if such skills were designed for families and kids. These results indicate that the parents were not aware of the skills with expletives that were *actually* developed for kids and published in Amazon's "Kids" category. In addition, about half of parents (44.2%) did not think the Sensitive skills were designed for families/kids, although these skills are actually in the "Kids" category on Amazon as well.

Parents' Comfort Level. We used a five-point Likert scale to measure parents' comfort levels if the presented conversations were between their children and Alexa. Figure 3.6 shows the participants' comfort levels for each skill category. These results indicate that parents were more uncomfortable with the Expletive skill conversations compared to the Sensitive skill conversations. In particular, 42.7% of the respondents expressed discomfort ("Extremely uncomfortable" and "Somewhat uncomfortable") with the Expletive skills, compared to only 12.1% with the Sensitive skills and 5.6% with the Non-Risky skills.

Chi-square tests show that parents' comfort with the Expletive conversations is significantly different from their comfort with the Sensitive conversations and from the Non-Risky conversations ($p < 0.05$). Some participants expressed their concerns about skills in the Expletive set by free-text responses, including "It doesn't seem appropriate to tell jokes like this to children (P148)," "Under no circumstances should anyone have a coversation [sic] with children about orgasms. This would be grounds for legal action (P163)," "I do not believe Alexa should be used in such a crass manner or to teach my child how to be crass (P210)," "Poop and poopy jokes don't happen in my household (P216)," and "It is too sexual (P123)." Beyond the skills shown in the survey, one respondent also recalled hearing similar skills such as "Roastmaster (P121)." Another respondent remembered something similar but was unable to provide the name of the skill: "We have asked Alexa to tell us a joke in front of our young son and Alexa has told a few jokes that were

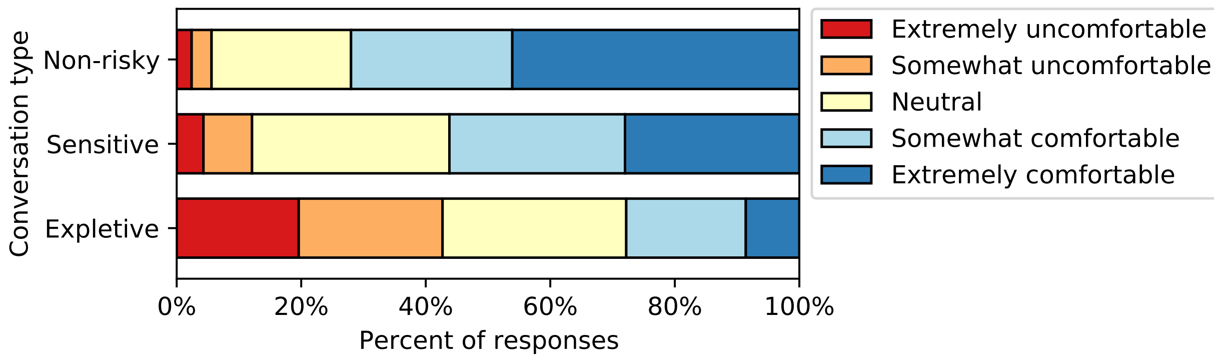


Figure 3.6: Participants' comfort levels if conversations of a particular type were to happen between the participants' children and Alexa.

borderline inappropriate (P140)."

We do not find any significant difference between parents' comfort with the Sensitive conversations versus the Non-Risky conversations. However, the Sensitive conversations involved skills asking for different types of personal information. Out of the 20 skills in the Sensitive set, 15 skills asked for the user's name, 3 asked for the user's age, and 2 asked for the user's birthday. We show the distribution of the participants' comfort level according to each type of personal information in Figure 3.7. This indicates that that parents expressed more discomfort ("Extremely uncomfortable" and "Somewhat uncomfortable") for skills that ask for the user's birthday (15.2% of respondents), compared with skills that ask for the user's name (11.8%) or age (11.5%). Some participants expressed their concerns about these skills by free-text responses, including "I don't like a skill or Alexa asking for PII (P115)," "I haven't had a similar experience but I think it is inappropriate for Alexa to be asking for the name of a child (P209)," "I don't know why it needs a name (P228)," and "I would not want Alexa to collect my children's imformation [sic] (P003)."

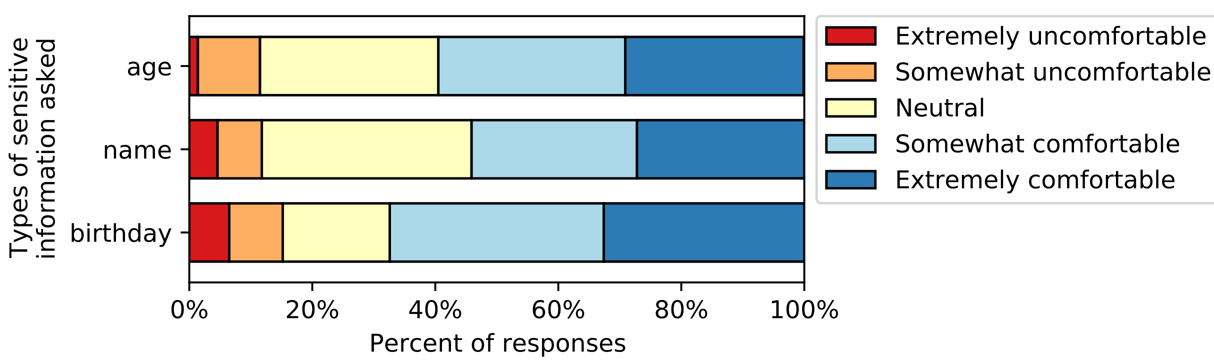


Figure 3.7: Participants' levels of comfort for each type of personal information, if the conversations happen between the participants' children and Alexa.

Amazon Echo Usage. Our results also show that most households with children use Echo devices other than the Echo Kids Edition. Echo Dot was the most popular type (46.4%) of Echo device in our participants' households. Only 27 participants (6.8%) bought an Echo Dot Kids Edition, which has parental control mode enabled by default. This shows that if children use an Echo, they likely have access to the types of Echo devices that do not have parental control mode enabled by default.

Furthermore, the majority of participants (91.8%) reported that their children do use Amazon Echo at home. Figure 3.8 shows the types of Echo that the participants own in their household associated with the breakdown of answers to the question "Do your kids use Amazon Echo at home?" Most parents allow their children to use Amazon Echo at home even without an Echo Dot Kids Edition. This indicates that many children have access to risky skills, as these skills can be used by default on Echo devices other than the Kids Edition.

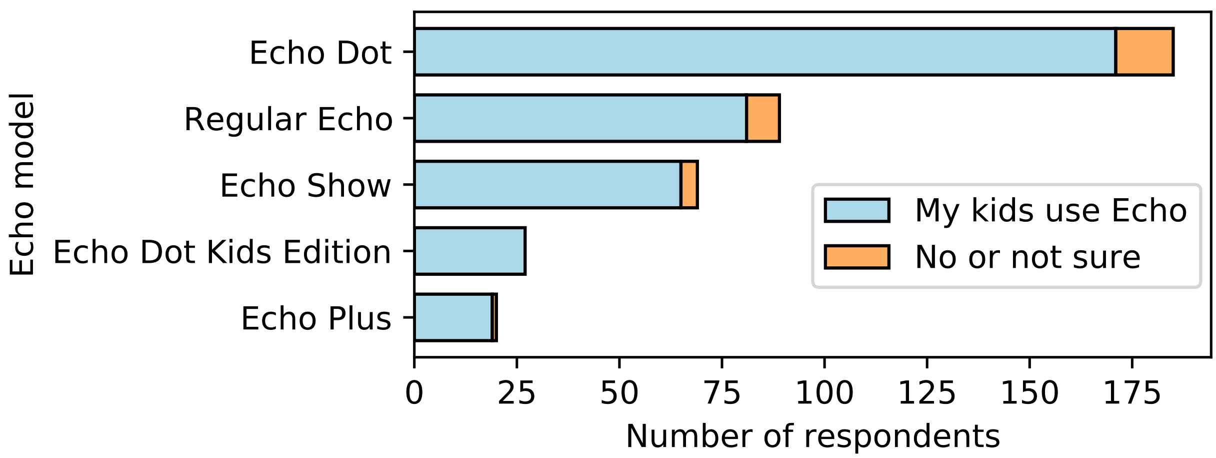


Figure 3.8: Types of Echo devices that the participants own in their households and the number of participants whose children use Amazon Echo at home. Echo Dot is the most popular device, and most households have children who use an Amazon Echo other than the Kids Edition.

Awareness of Parental Control Features. We analyzed the responses to the question: "Does Amazon Echo support parental control?" In total, 76.3% participants said "yes," 0.4% said "no," and 23.3% were unsure. For participants who had an Echo Kids Edition, almost all (92.6%) said "yes," 7.4% said "no," and none was unsure. In contrast, for participants without an Echo Kids Edition, only 74.1% said "yes," and 25.4% were unsure. This indicates parents who did not buy the Echo Kids Edition are less likely to know about the parental control feature.

For those who said "yes" (i.e., they knew Amazon Echo supports parental control), we further asked if they used parental control. In total, only 29.4% used parental control. Specifically, 64.0% of participants that had Echo Kids Edition said they used parental control, but only 23.7% of those who did not have Echo Kids Edition used parental control. Given that many fewer participants had an Echo Kids Edition (only 27 people out of 232), the majority of parents did not use parental control for their Echo devices at home. This result again indicates that children are more likely to have access to risky skills.

Furthermore, although some participants reported that they used parental control, many of them did not really know what the feature involved. For participants that said they used parental control, we asked them to tell us the name of the parental control feature. As long as their answer contained “free” and “time” (the correct answer is “FreeTime”), we considered their answer correct. 66.7% of participants that had an Echo Kids Edition were able to correctly name the parental control feature, but only 42.3% of those that did not have an Echo Kids Edition could do so.

Takeaway. Our results show that parents express significantly more disbelief and concern about conversations from Expletive skills versus Sensitive skills. This is worrisome, because all skills displayed on the survey are real and can be invoked by *anyone* now, including children. The fact that parents’ responses to Sensitive skills were not significantly different from their responses to Non-Risky skills also illustrates a potential lack of understanding that skills are developed by third parties and may pose a privacy risk.

3.8 Confounding Utterances Analysis

To understand how easy it is for kids to accidentally trigger skills (**RQ3**), we performed a systematic analysis to identify confounding utterances and corresponding skills. We identified the set of potential confounding utterances from the skills’ info pages, each of which might invoke multiple different skills, and used SkillBot to collect and analyze conversations started by these utterances.

3.8.1 Discovering Confounding Utterances

To identify confounding utterances, we first created a dictionary with utterances extracted from Alexa skill info pages as keys and lists of skills that respond to each utterance as values. For each utterance, we removed punctuation and used lowercase to keep the format consistent. We filtered the dictionary to contain only the confounding utterances corresponding to at least two different skills. We then wanted to discover which skill Alexa chooses to enable/invoke when given a confounding utterance that is shared between multiple skills. When Alexa prioritizes a skill which does not match the user’s intent, it poses a potential risk to the user. This is especially true if the user is a child. For example, if Alexa prioritizes a non-kid skill over a kid skill that shares the same utterance, a child user could inadvertently gain access to risky content.

We discovered a set of 4,487 confounding utterances, each of which was shared between two or more skills. Of these 4,487 utterances, 110 (2.5%) belonged only to kid skills, 581 (12.9%) belonged to both kid and non-kid skills, and 3,796 (84.6%) belonged only to non-kid skills. We defined these three categories of confounding utterances as "Kids Only," "Joint," and "Non-kids Only," respectively. We further identified utterances belonging to skills with the same skill name and skill icon—properties that would make it difficult for users to distinguish among these skills even by visually inspecting the skills' webpages. For confounding utterances in the Kids Only category, we found that 6 (5.5%) out of 110 were shared among skills with the same name and icon. For those in the Joint category, we found that 48 (8.3%) out of 581 were shared among skills with the same name and icon. For those in the Non-kids Only category, we found that 577 (15.2%) out of 3,796 were shared among skills with the same name and icon.

3.8.2 Testing Confounding Utterances

We used SkillBot to test the confounding utterances identified in the discovery step. For each utterance, we disabled all skills and entered the utterance. We then checked if any of the skills that shared the utterance had been enabled (Table 3.5).

	Kids Only (N=110)	Joint (N=581)	Non-kids Only (N=3,796)	Total
Invoked irrelevant skill	64	367	1,999	2,430
Invoked relevant skill	46	57	1,797	1,900
Invoked relevant skill but prioritized non-kid skill	-	157	-	157

Table 3.5: Number of confounding utterances of each type and corresponding behaviors.

Kids Only: We found that 64 (58.2%) out of the 110 confounding utterances in this category invoked an irrelevant skill that was not in the list of skills associated with the utterance itself. The remaining 46 utterances (41.8%) invoked a relevant skill within the list of associated skills. Examples include the utterances "Start owl facts" (shared by two different kid skills) and "Open animal sound quiz" (shared by three different kid skills).

Both Kids and Non-kids (Joint): We found that 367 (63.2%) out of the 581 confounding utterances in this category invoked an irrelevant skill that was not in the list of skills associated with the utterance itself. The remaining 214 utterances (36.8%) invoked a relevant skill within the list of associated skills. However, 157 out of these 214 utterances (73.4%) prioritized invoking a non-kid skill over a kid skill. For example, "Start human body quiz" is a confounding utterance that is shared between two skills. One of them is a skill published for kids. The other one is a general game skill which is marked as having mature content.

Non-kids Only: We found that 1,999 (52.7%) out of the 3,796 confounding utterances in this category invoked an irrelevant skill that was not in the list of skills associated with the utterance itself. The remaining 1,797 utterances (47.3%) invoked a relevant skill within the list of associated skills. Examples include "Start movie picker" (shared by five different skills) and "Give me a random fact" (shared by seven different skills).

Takeaway: It is risky if a confounding utterance is shared between a kid skill and a non-kid skill. Our analysis shows that children can accidentally invoke a non-kid skill while trying to use a kid skill. An adversary can exploit this problem to get child users to invoke risky non-kid skills.

3.8.3 Confounding Utterances' Risks to Kids

Our confounding utterance analysis shows that children can accidentally invoke a non-kid skill while trying to use a kid skill. Furthermore, confounding utterances could be an attack vector for an adversary to expose children to risky content without having to publish a kid skill, since non-kid skills do not need to follow the content requirements in place for kid skills. We discuss two risks that we observed during our confounding utterance analysis as follows.

Non-kid skills risks To gain more insights into the potential risks from non-kid skills, we sampled a set of non-kid skills containing 50 skills from each of the other 22 categories (1,100 skills in total). We ran SkillBot to interact with each skill and record the conversations. We collected 7,356 conversations from 1,073 non-kid skills after removing 27 skills that resulted in errors or crashes. We used the same approach as our kid skill analysis presented in Section 3.6.2 to analyze this sample. As a result, we identified 4 skills with inappropriate content and 16 skills asking for personal information, such as name, age, birthday, phone number, zipcode, and address. Of the 16 skills asking for personal information, 8 skills did not request any permissions from the user. We further investigated the 3 skills that requested permissions. Only one of them requested the proper permission for the personal information it queried.

Sneaky skills risks During our analysis, we noticed an abnormal behavior of the Alexa skills interface. In particular, through the "Your skills" tab on both Alexa web interface and mobile app, users can view a list of skills that are enabled in their account. We observed that the interface sometimes showed an empty list of skills although we actually had some skills enabled. We managed to further investigate this issue by enabling/disabling the skills via API calls and asking Alexa about the status of the skills to verify. We found that the skills were actually enabled and disabled properly in the backend, but the front-end did not display their status properly. From our user study, some participants mentioned in free-text responses that they preferred to monitor enabled skills on their own. Additionally, our skill analysis found that in some cases users have limited or no ability to check which skill they actually invoked due to skills having similar names and icons. Therefore, this front-end bug could become a security issue as it increases the stealthiness of malicious skills. In particular, a malicious skill can get accidentally enabled and remain invisible to users. In one possible attack, an adversary could craft a malicious skill that exploits a confounding utterance to become unexpectedly enabled/invoked by the user, causing the skill to become "sneaky," as the front-end bug prevents the user from viewing what skills are enabled.

3.9 Discussion and Future Work

This section provides some suggestions for building safe VPA for children. We also acknowledge some limitations of our study and discuss the potential future work.

Suggestions for Building Safe VPA for Children. Our study shows that the current Alexa skill vetting process is not effective, even for skills targeting children – a vulnerable population. Although there are stricter requirements for developing and publishing kid skills, risky kid skills still exist on the store. This means VPA service providers need a more robust vetting system to ensure that the published skills adhere to policy requirements. Furthermore, since skills can be hosted on third-party servers, it is hard for VPA service providers to control what happens in the backend. A nefarious skill developer could easily manipulate the backend to turn a benign skill malicious. Thus, a continuous vetting process is important to ensure consistent adherence to policy requirements.

VPA service providers also need to improve detection and limitation of confounding utterances, especially those that may unintentionally invoke different skills than the user intended. Without further protection, the existence of confounding utterances enables children to accidentally invoke non-kid skills, potentially exposing them to inappropriate content. Third-party skill designers could be required to register invocation phrases upon posting their skills on VPA provider hosting platforms. Like email addresses or domain names, preventing overlap of skill invocation phrases would keep children and other users from accidentally opening an unwanted skill with potentially risky content.

Our user study results showed that many parents do not know about or do not use parental control features on VPA devices. This likely means that more user-friendly parental control features are needed to reduce burden on parents while providing strong protections for child users. Parents should be encouraged to use parental control features for VPAs, especially on devices placed in a shared space. Existing recommendations for the design of parental control software in other domains likely carry over to the VPA space and would help parents' prevent children from accessing risky non-kid skills.

Limitations. The Alexa developer console employed by Skillbot has some limitations [100] as compared to physical Echo devices. The limitation most relevant to our study is the inability to collect the content of audio playbacks from skills (e.g., music skills). Thus, we did not consider such skills in our dataset. Since audio playbacks are rare (as shown in Section 3.5.3), we believe that this limitation is a reasonable trade-off for Skillbot's ability to efficiently perform analysis at scale and does not undermine our results.

Our user study's results were based on self-reported data, which means the responses might be influenced by social desirability. Specifically, people are often biased in self-reports toward social norms [104]. To mitigate this, we tried to use neutral wording for our questions. As with any self-reported surveys, participants may choose the first answer that satisfies without thinking carefully about the question. Thus, we included attention check questions to filter out such inattentive participants from our study. In addition, our user study protocol includes incentives for survey completion, which might cause a bias. We also reviewed the survey responses and filtered out invalid responses as described in Section 3.7.4. We believe that if there are any remaining data quality issues, they are minor and do not affect our findings.

Future Work. We maintain a dataset of recorded conversations with Alexa skills for future research. This dataset will increase in size over time as we run Skillbot on new skills and longitudinally to explore new conversation branches of existing skills. Future research using this dataset could provide insights into new and ongoing risky skill behaviors and help with the creation of rules and systems to protect users.

While this study focuses on risky content in kid skills, we show that children can potentially be exposed to non-kid skills through confounding utterances and a lack of parental controls. Thus, future work could further investigate the risks posed by non-kid skills that can be accidentally invoked by children.

Finally, we focused our analysis on Alexa, but it would be easy to extend Skillbot to work with other VPA platforms in order to investigate risky skills on those platforms. Skillbot could also be integrated into a public service for consumers to vet "black-box" skills before installing them (i.e., a potential countermeasure for consumers).

3.10 Conclusion

We contributed an automated skill interaction system called SkillBot, analyzing 3,434 Alexa kid skills. Our system identified a number of risky skills with inappropriate content or personal data requests and described a novel confounding utterance threat. To further evaluate the impacts of these risky skills on children, we conducted a user study of 232 U.S. parents who use Alexa in their households. We found widespread concerns about the contents of these skills, combined with general disbelief that these skills might actually be available to children and a low rate of adoption of parental control features.

Chapter 4

Automated Risk Analysis and Voice-controlled App Monitoring

4.1 Overview & Goal

The previous chapter showed how we built a system to analyze the voice apps by automatically interacting with the apps, addressing the challenge of no binary installation and source code. Our system identified several risky child-directed apps that asked for personal information or contained inappropriate content. However, it is unclear how such apps passed the service provider (Amazon Alexa)'s vetting process. Therefore, in this chapter, our goal was to find the reasons behind policy-violating skills remaining published in the wild. We revealed that, in addition to vetting, Amazon also deploys a skill monitoring scheme to continuously check the skills' behaviors. However, this monitoring scheme also had flaws. We reverse-engineered this monitoring process and identified the limitations. With such limitations, there could be policy-violating skills that passed the vetting process and remained published for the users. We aimed to protect user privacy when they interact with such skills.

4.2 Introduction

The global intelligent virtual assistant market was valued at billions of dollars and is expected to continuously grow from 2021 to 2028 [105]. As smart software agents that can provide services based on user commands or questions, virtual personal assistants (VPA) like Amazon Alexa and Google Assistant are becoming increasingly popular with families around the world.

In this study, we perform a large-scale analysis of Alexa skills collected in the US skill store and a systematic study of the Alexa skill ecosystem, especially the risks caused by the developers' ability to dynamically change a skill's back-end code. We find that the vetting policy enforced by Amazon is following different criteria from what is claimed in developer requirements. We uncover the skill monitoring process which has not been studied before. Besides, Amazon allows developers to deploy a skill's back-end logic to a third-party server that is not under Amazon's control. This is very risky because malicious developers can submit a benign skill and change its back-end behavior logic after the skill is published in the skill store. In this way, it is able to circumvent the whole vetting process and permission model easily. A skill verified as a benign one may request user personal information (e.g., phone number) verbally. We call this risk *skill's behavior change*. When a skill's behavior change happens, users may be credulous and just reveal personal information. We also find some evidence showing such malicious skills do exist in the wild. Even though Amazon strengthened skill vetting by appending a new monitoring process to the initial vetting process, we find it is still insufficient to defend against a skill back-end change. Our test skills successfully pass the vetting process and perform malicious behavior changes without being detected. To help protect user privacy, we propose an effective real-time skill behavior monitoring system that detects suspicious skills and notifies the user during interactions.

Skills deployed on external servers can introduce privacy risks to users [2] because malicious developers can change their back-end code anytime after skill approvals. However, there is no prior research about skill behavior monitoring approaches to detect malicious behavior changes, mainly due to the following challenges.

Challenges. First, the skill response, or the *behavior* of skill, should always be consistent with its functionalities. When focusing on the suspicious skill response asking for personal information, we need to distinguish the information request (e.g., "What's your phone number") from other responses (e.g., "Do you want to continue?"). This needs a corresponding natural language processing method. In addition, the most challenging part is to analyze the skill description, which describes the skill functionalities. The variability of semantic expression makes skill descriptions diverse even when two skills are doing the same thing. Straightforward keyword-based approaches can only handle those skill descriptions with matched keywords. It is a challenge to develop a semantic interpretation approach to handle the ambiguity of natural language and find the correlation between skill response and skill description. Second, when developers put the skill's source code in an external server, the skill behavior logic (how the skill responds to user utterances) becomes a black box. Moreover, malicious developers can change the back-end code anytime to request private information from users. This issue can only be solved by a real-time skill behavior monitoring system that runs in the background whenever the user communicates with skills. However, a real-time running system must have an unnoticeable delay, which requires the skill monitoring system to have high accuracy and low latency.

Contributions. This study has the following contributions:

- We present a comprehensive understanding of how Amazon actually monitors the skills submitted for their Alexa platform (Section 4.4). We comprehensively analyze the risk of the skill's back-end change. Compared to previous work, we present more approaches to perform skill back-end changes that bypass Amazon's monitoring process.
- Amazon's current vetting mechanisms failed to protect users against skill behavior change. We propose a skill behavior monitoring approach working at run-time when the user is interacting with the skills. No matter how malicious developers design their skill's dynamic logic to evade Amazon's monitoring, once the skill asks for the user's personal information, it will go through our monitoring system's check at run-time which then informs the user of the risks. Our proposed system performs semantic understanding of the text by applying textual entailment and achieves negligible overhead by hashing the pre-computed results (Section 4.5).
- We conduct a measurement study of 54,587 published Alexa skills using our skill behavior check approach (Section 4.5.6). Of 637 skills that requested personal information from users, we identified 141 suspicious skills that asked for personal information without proper permissions or the use of such information for their functionalities.

Key Findings. Our key findings in this study include:

- We discover the monitoring process after a skill gets published and reveal the criteria Amazon follows to identify a violating skill, which has not been studied before (Section 4.4.1 and Section 4.4.3).
- We discover a new vulnerability called "versatile intent" that allows adversaries to stealthily collect any type of information by manipulating their published legitimate skill (Section 4.4.2).
- We identify the necessity for a defense system at run-time. In particular, the service provider does not have full control over the published skills, especially those hosted on third-party servers. The behavior changes of such skills are unpredictable. Our experiments showed that the skills could adopt certain patterns for their behaviors to bypass the vetting (Section 4.4.4).
- Privacy-invasive skills are still an underlying threat (Section 4.5.6). Of 637 skills that requested personal information from users, we identified 141 suspicious skills without proper permissions or functionalities.

4.3 Related Work

Virtual personal assistants (VPA) have been studied as one category among general IoT devices in some prior works [6, 7, 8]. But lately, virtual personal assistant devices have drawn attention as a popular category of IoT equipment. Many

Table 4.1: Comparison with related work. ○○: not analyzed, ●○: partially analyzed, ●●: analyzed.

Metric	Zhang et al.[110]	Cheng et al.[18]	Guo et al.[69]	Lentzsch et al.[2]	Liao et al.[17]	Our work
Back-end Change	● ○	○ ○	○ ○	● ○	○ ○	● ●
Description Analysis	○ ○	○ ○	● ○	○ ○	● ○	● ●
Vetting Criteria	● ○	● ○	● ○	● ○	○ ○	● ●
Monitoring Process	○ ○	○ ○	○ ○	○ ○	○ ○	● ●
Run-time Defense	○ ○	○ ○	○ ○	○ ○	○ ○	● ●

researchers put effort into how to attack or fool the system. There are some works focusing on adversarial attacks against speech recognition [9, 10, 11, 12]. Some researchers focused on detections or defenses [15, 16] against such voice recognition attacks.

The speech recognition system, which is the core of the voice assistants, was known to have misinterpretation vulnerability [3, 106, 107, 108]. Exploiting the misinterpretation problem, an adversary could impersonate the voice assistant system or other skills to eavesdrop on users. Researchers in SR Labs built malicious skills [109] on Alexa to ask for users' passwords or eavesdrop on users' conversations. Apart from the speech recognition component, prior work [110] also analyzed and evaluated the security of the succeeding Natural Language Understanding (NLU) component.

Smart speakers (i.e., Alexa products/google home) work more like a black box to users. Based on some previous works [21, 22], users have an incomplete understanding of the smart speaker model and are concerned about their privacy [23]. Recently, researchers put more effort into reverse-engineering the smart speakers' vetting mechanism[18, 19]. Cheng et al. [18] built plenty of test skills to disclose the skill certification process on VPA platforms. Our work is different from Cheng's study [18]. We discovered that the certification process is not the only vetting mechanism on the VPA platform. After skills get published on the market, there is a repeated vetting process called the monitoring process, which will run periodically to monitor skill behavior. We reverse-engineered this unstudied process and provided a more comprehensive understanding of the VPA vetting mechanism.

Comparison with prior work: In this study, we present a systematic analysis of the Alexa skill ecosystem focusing on skill's back-end change and disclose the procedure and limitations of the Amazon's monitoring process. We highlight that even with Amazon's monitoring process for the published skills, skill's back-end change by exploiting "versatile" intents or dormant intents is still a risk to user privacy security. Table 4.1 gives an overview of how our work differs from related studies. To better protect the user from deceptive or misleading information, we propose a real-time skill behavior monitoring mechanism (real-time defense) that has not been demonstrated before. Zhang et al. [110] only discussed how the developers replace back-end audio files. Lentzsch et al. [2] illustrated how to exploit dormant intents to manipulate the back-end code logic of the skills. In our work, we study the back-end code change risks more thoroughly with more test skills. In particular, we not only study the dormant intent but also discover a new approach called "versatile intent" that is harder to detect. Previous studies [18, 69, 111] identified policy-violating skills and found that Amazon conducted a vetting process differently from what is described in its documentation. However, they did not explore the behind-the-scene criteria that the vetting process follows. We explore and present such criteria in Section 4.4.1. Besides, we uncover that there is not only the vetting process but also a monitoring process after the skills get published. We investigate both the vetting process and the monitoring process to provide a more comprehensive understanding of how Amazon manages the skill ecosystem (Section 4.4.1 & Section 4.4.3). Liao et al. [17] analyzed the consistency between the privacy policies provided by the skills and the corresponding skill descriptions. However, skill description might not cover everything the skill actually does and privacy policy is not a reliable criterion to measure the risks because privacy policy is not taken seriously by users, developers, and Amazon. In contrast, we focus on the actual behaviors of the skills. While prior work from Guo et al. [69] performed skill description and skill behavior analysis, they employed a manual analysis approach on 100 selected skills. In our work, we present an automated approach to analyze the consistency between skill functionality and description, which is a foundation of our real-time skill monitoring system (Section 4.5). Lastly, we identify the need of protecting users against suspicious skills in the wild and build the first real-time skill behavior monitoring system to protect user privacy against malicious skills (Section 4.5).

4.4 Systematic Study of Alexa Skill Vetting and Privacy Risks

In this section, we shed more light on the Alexa skill vetting mechanisms. First, we uncover Amazon's criteria for vetting (Section 4.4.1). Second, previous research mostly studied the skill certification process that happens before the skill gets published [18, 19]. We discover that the skill certification process is just the first stage of the whole vetting pipeline. We take a further step to uncover the skill monitoring process happening after a skill gets published (Section 4.4.3). Figure 4.1 illustrates an overview of the skill development and vetting process. Then, we demonstrate approaches that can bypass the whole vetting pipeline to perform malicious skill's behavior changes (Section 4.4.4).

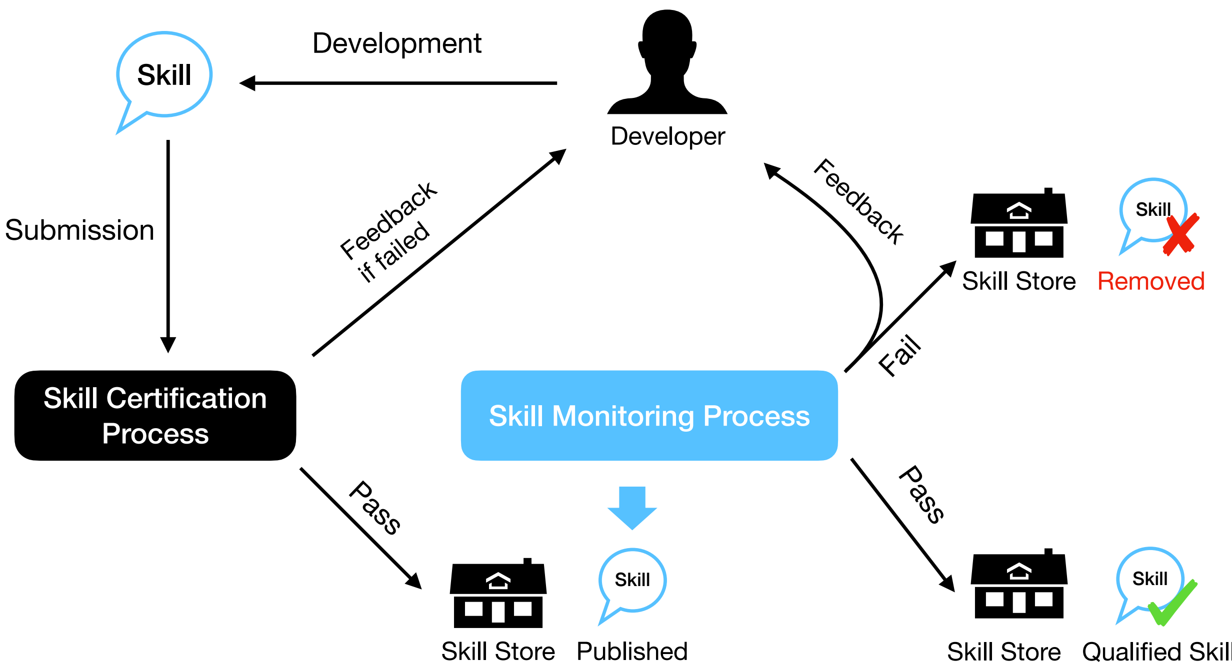


Figure 4.1: The full vetting/monitoring process of Amazon Alexa

4.4.1 Demystifying Amazon's Skill Vetting

Finding I: Previous studies found skills violating privacy policy on the market or asking for personal information without declaring permissions [17, 18, 69], which was concluded to be the limitation of Amazon's vetting. We further uncover how Amazon's vetting actually works to better understand why such violating skills were allowed to be published. In short, it follows a criterion: functionality correlation. The user information requested by skill is not necessarily declared in the permission list. The skill can ask for the user's personal information through voice interaction as long as it declares such information is necessary for the skill's functionalities.

Alexa developer documentation regulates how the skills request user information. In summary, it requires the developer to provide a permission list and describe the skill's functionalities to the users if the skill requests any kind of personal information. The detailed requirements are presented in Appendix B.1. Although the documented vetting policies have detailed and strict requirements, the actual vetting process works differently. For example, many published skills did not provide a privacy policy when they were supposed to. Prior studies [17, 18] presented such measurements. However, details about how Amazon's vetting process actually works are still underexplored.

How Amazon enforces the policies

Cheng et al. [18] showed that some skills asking for a user name in the first interaction could still pass the certification process even though they did not claim any permission. If Amazon followed the vetting criteria precisely the same as its documentation, they would not have missed the violations in the first interaction with the malicious skills. What are the tricks for these skills to bypass Amazon's vetting? It is very likely that Amazon is enforcing vetting criteria differently from what is claimed in the documentation. Previous works attributed it to Amazon's leniency, which oversimplified the issue. We uncovered and validated that "leniency" follows some rules. To explore this issue, we built some test skills to reverse engineer the vetting process.

If a skill requests personal information, the skill needs to provide the following three items according to the developer documentation: the description to depict skill functionalities, a permission list, and a privacy policy disclosing the requested information. We designed some test skills to make it clear which parts are not considered seriously by Alexa vetting through the controlled variable method. We first built our test skill *Mascot Box* without a permission list (it still had a privacy policy and skill description). The skill's functionality is to provide some sweet words to the user when invoked. However, it will also ask for the user's phone number in the first response. As a result, it failed the certification process of which Amazon's feedback said *"... Your skill is requesting information that is not relevant to the skill's functionality. Namely, your skill request phone number."* The feedback inspired us to think that an important vetting criterion could be the correlation between functionality and actual behavior. So we validated this criterion by re-making a test skill to first ask for *full name* and then give a greeting response "Hello, *A*" after the user provides his/her name *A*. This time, the test skill successfully passed the certification process. To further confirm the discovered criterion for other types of personal information, we built similar test skills that asked for a phone number, email, and location. Test skills include test skill *Sweet Text* which will send a sweet message to the user's phone number or email, test skill *My Weather* which provides local weather to the user, etc. They all became qualified skills and were published on the store.

The above experiment shows that the skills are allowed to collect customer contact information if the request information has a correlation with the skill functionalities, which can explain why some aforementioned test skills violating privacy requirements could still pass the certification process in prior works. For example, the test game skills [18] asking for a name in the first response passed the certification process because the Alexa vetting team considered a user name could improve user experience (i.e., it has a correlation with the skill functionality to some degree). Specifically, regarding *name* information, we found any game skills were allowed to ask *"What's your name?"* as long as a greeting *"Hello, [name X]"* followed after the user provided a name.

4.4.2 Manipulating Skill Behaviors at Run-time

Finding IV: Malicious developers can change the back-end logic of skills deployed on the third-party server after the skills get published on the store to have users disclose personal information. The back-end change can be achieved by dormant intents and "versatile" intents. The latter approach is novel and stealthier, motivating the need for a run-time monitoring approach while the user is interacting with the skills.

When a skill is deployed on a third-party server where Amazon can not access its back-end code, developers can arbitrarily change the skill's behavior logic at any time. This makes the skill able to dynamically change its behavior after it gets published on the store, e.g., the skill can ask for a user phone number that has no correlation with its functionalities. Credulous users may just reveal their information when they are requested. Currently, there is no vetting mechanism that is able to detect such suspicious behavior changes.

Malicious developers have two approaches to exploit the skill's behavior changes for their published skills. The first approach is to leverage "dormant" intents, which was previously discussed by Lentzsch et al. [2]. In our work, we discovered a new approach that leveraged "versatile" intents. We describe the two approaches and provide the comparisons as follows.

Dormant intent

A malicious developer can craft a safe skill with an unused intent (so-called "dormant intent") that collects certain sensitive information (e.g., phone number). The vetting process will not find any suspicious behavior of the submitted skill because the trigger logic does not get designed for the dormant intent yet. However, after the skill gets published, the developer can change the back-end code to activate the dormant intent. Whenever users trigger that intent, the skill will ask the users for the sensitive information that the intent was crafted for. Figure 4.2 shows how a skill that originally does not request any personal information gets its dormant intent activated after passing the certification process.

Versatile intent

Amazon maintains a list of slot types that define how phrases in utterances are recognized and handled for intents, such as *AMAZON.PhoneNumber, AMAZON.City, AMAZON.Color,* etc. No matter which kind of information a skill wants to obtain on the back end, it has to choose a corresponding intent slot type for its purpose. There are two categories of slot types: *List* and *Numbers/Dates/Times*. *List* includes slot types that represent a list of items (text), while *Numbers/Dates/Times* includes slot types that convert the user's utterance into data types such as numbers and dates. Amazon tried to provide separate slot types for different types of information to prevent developers from requesting more than what they need [1]. However, malicious developers can use an intent designed for information *A* to collect

[1] https://developer.amazon.com/en-US/docs/alexa/custom-skills/slot-type-reference.html

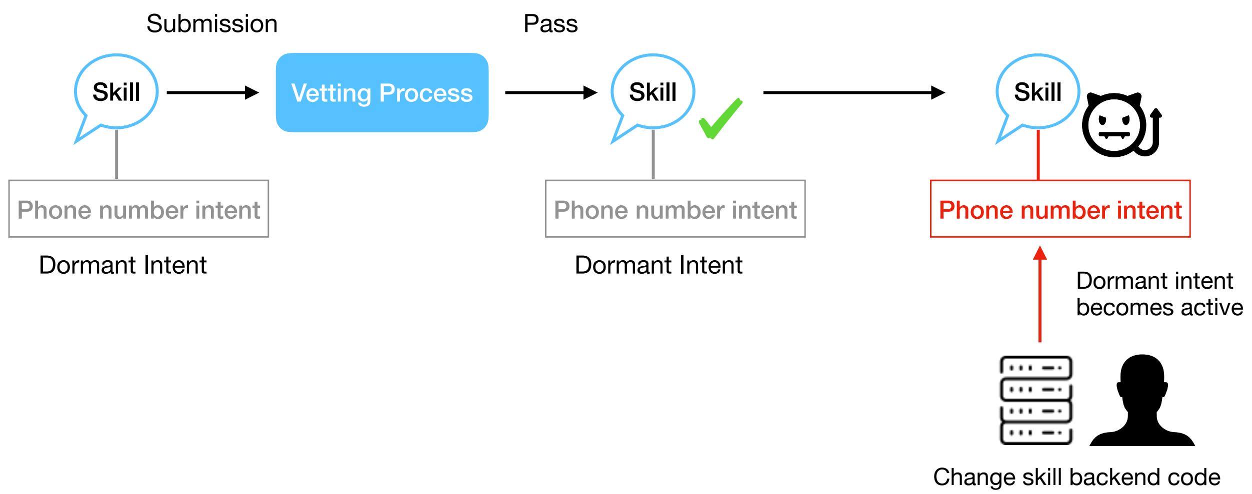


Figure 4.2: Overview of how a "safe" skill passes the vetting process then activates its dormant intent for the usage of collecting sensitive information

information *B*; we call such intent "versatile intent". For example, *AMAZON.Number* can be applied to collect phone numbers, zip codes, or other number-related information. Figure 4.3 shows our proof-of-concept test skill that uses *AMAZON.number* to collect phone numbers without properly registering a dormant intent with *AMAZON.PhoneNumber*. Our test skill, named *"Guess Number"*, is a game skill that generates a random integer number and asks the user to guess it. The *AMAZON.Number* intent is necessary for its basic functionality, but a malicious developer can secretly leverage it to collect personal information such as phone numbers through back-end change. Similarly, there are many other versatile intent slot types that can be exploited. For example, *City* types can be used to collect user names (see Figure B.1 in Appendix B.3). Interestingly, we identified several versatile intent slot types that are non-restricted (Table 4.2). Different from regular slot types, these slot types can record any text or number inputs. For example, *AMAZON.StreetName* can be used to collect any information other than the street name. Figure 4.4 shows a demo of our test skill exploiting *AMAZON.StreetName* to collect the user's email address.

Table 4.2: Non-restricted intent slot types provided by Amazon Alexa. These slot types have blank descriptions and can be exploited to collect any information.

Category	**Non-restricted Slot Types**
List (Text)	AMAZON.Anaphor AMAZON.RelativePosition AMAZON.StreetName AMAZON.VisualModeTrigger
Numbers/Dates/Times	AMAZON.Ordinal AMAZON.PhoneNumber

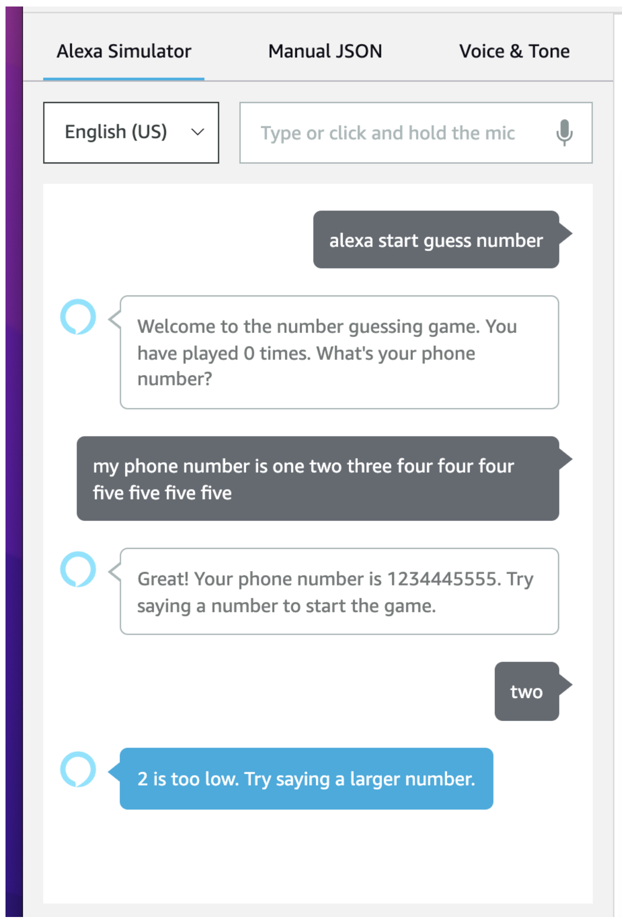


Figure 4.3: "Guess Number" skill leveraging *AMAZON.Number* instead of *AMAZON.PhoneNumber* to collect the user's mobile number.

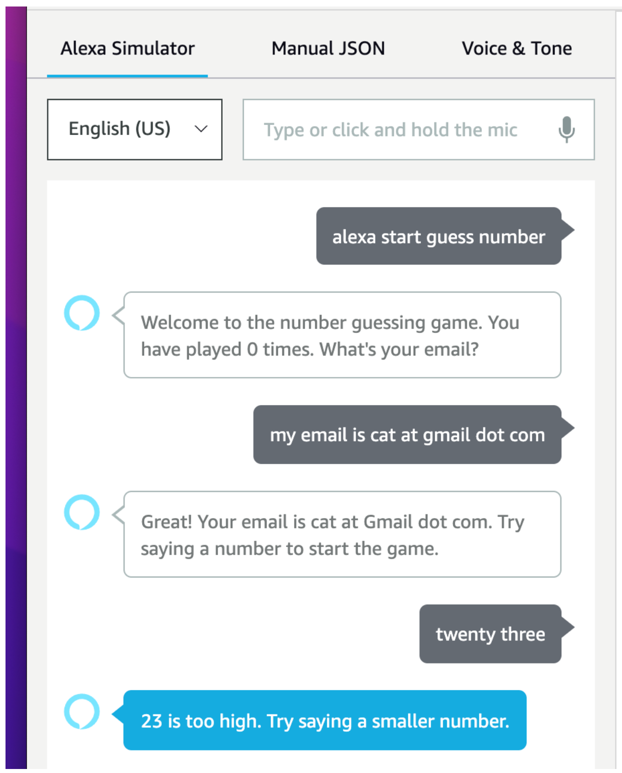


Figure 4.4: Leveraging *AMAZON.StreetName* to collect the user's email address. *AMAZON.StreetName* can be employed to collect any kind of personal information.

Versatile vs Dormant

Versatile intent exploit has not been discussed before. Versatile intents are more concealed loopholes compared to dormant intents. Dormant intents are unused intents that could be easily detected during the vetting phase. Amazon can check all intents crafted in the skill and identify unused intents that can be dormant intents. However, it is different for versatile intents. Versatile intents are intents originally used for legitimate skill functionalities (satisfying vetting requirements) but can be exploited for a different purpose at run-time. In this case, run-time monitoring while the user is interacting with a skill is needed. As shown in Figure 4.3, an adversary can build a skill that necessarily has certain versatile intents serving its functionality and then later exploit them to collect different personal information. There is no way for Amazon to foresee whether or when the skill's back-end logic would be modified for malicious purposes.

4.4.3 Reverse Engineering Amazon's Skill Monitoring Process

Finding II: The so-called monitoring process is an underexplored process in the skill vetting mechanism. It intermittently tests the behavior of recently certified skills in the store during a certain period. Suspicious skill behavior will cause skill removal. However, it is a periodic testing process instead of a supposedly continuous monitoring process and can be bypassed easily because it follows certain patterns.

Prior studies [2, 18] put a lot of effort into figuring out how Amazon's skill certification process works and its loopholes. However, that is not the whole picture of the vetting process. Even though a skill passes the certification process, the vetting is not finished yet. Amazon repeatedly tests published skills for a period, which is called the "monitoring process," as shown in Figure 4.1. If the skill shows suspicious behavior, i.e., asking for personal information which has no correlation with its functionalities, it will get removed from the store. When our test skill passed the certification process, we changed its back-end logic to ask for personal information unreasonably. After several days, we received feedback informing us that the skill did not pass the "monitoring process" and was removed from the store. The feedback email is displayed in Figure 4.5 where we highlighted the name of this process and the reason why our test skill was removed. To the best of our knowledge, this monitoring process has not been studied before.

To understand how this monitoring process actually works, we built and submitted multiple test skills of different functionalities to the Alexa platform. On the skill back end, we logged the requests made to each skill. We illustrate our analysis and findings in the following paragraphs of this section.

Misleadingly named "monitoring process"

Amazon claimed that they deployed a monitoring process that continuously monitors and tests the skills on their market, as shown in Figure 4.5. To understand how this monitoring process works, we first published three safe skills. Once they got published in the store, we changed the back-end logic of our skills to ask for personal information unreasonably and not disclose it to the user. All those three skills were not removed from the store until seven days later. Our experiment suggests that Amazon's monitoring is actually not a continuous process, which could leave an attack window for around seven days after the skill gets published.

amazon alexa

Dear inspx,

In order to ensure that our skills are meeting the high quality bar expected by our customers, we continuously monitor and test live skills in our catalog. Your skill 'Mascot Box' was tested as part of this monitoring process and has failed certification. Until the issue noted below have been resolved, your skill has been removed from the skills catalog and will no longer be available for use by customers. The detailed certification results are compiled below.

Your skill is requesting information that doesn't appear to be relevant to the skill's functionality. Namely, your skill collects phone number which is not required as part of skills functionality. Please only request personal information from a customer when it is necessary for an element of your skill's functionality.

Steps to Reproduce:

User: "Alexa, open mascot box"

Skill: "What's your phone number?" & session remains open.

Please address these issues and resubmit the skill at your earliest convenience to provide a better customer experience.

Figure 4.5: Skill monitoring process feedback from the Amazon Alexa platform

Functionality consistency-based monitoring

As illustrated in Section 4.4.1, whether the information request from the skill has a correlation with its functionalities determines if the skill is able to pass the Amazon vetting mechanism. With our test skills, we validated that the criterion is a generic standard applied to all information permissions, including name, postal address, phone number, email, and zip code.

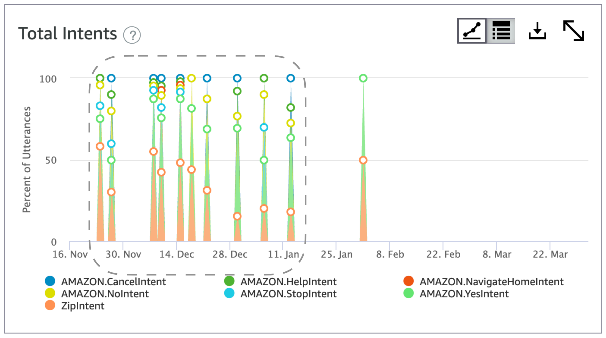
We first had five skills published in the store, then changed their back-end code to request those above five different types of personal information without reasons and corresponding permission requests. The skills were intentionally modified to ask for user information in the first response whenever launched so that Amazon's monitoring process would not miss these suspicious requests. All skills were removed from the store within around one week, just like the previously published three test skills. Next, we published another five test skills, and again, we deliberately changed the back-end logic to have them ask for the five kinds of information, respectively. But the different part in this round was that those information requests have a correlation with the skill's functionalities. As a result, 5/5 skills survived the monitoring process this time even though they did not have corresponding permissions. This experiment validated that the vetting criterion for suspicious skill determination is based on whether the requested information is consistent with the skill functionalities.

Periodic monitoring pattern

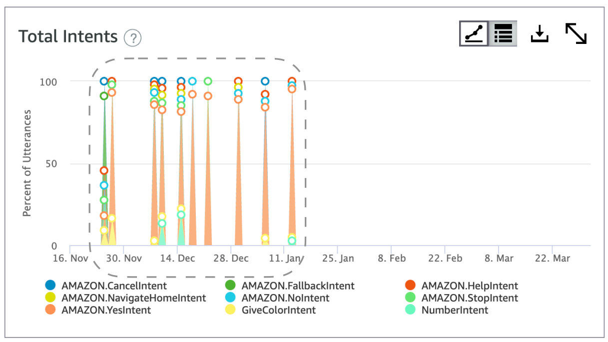
Since we found that Amazon's skill monitoring is not a continuous process, we wanted to further discern the pattern. To do this, we collected backend logs from our test skills, which were subjected to Amazon's test queries. However, Amazon anonymizes these test requests, making them indistinguishable from genuine user interactions. The concurrent use of our test skills by Amazon's vetting process and potentially real users complicated the process of identifying Amazon's monitoring patterns. To address this challenge, we undertook a retrospective examination of the activity logs from our test skills. We deduced Amazon's request pattern based on the types of requests made and the timing of our skill deactivation. Figure 4.6 shows the skill activity visualization metrics of our three test skills published on the same day. The first round of requests was logged on November 20th, approximately seven days post-publication, signifying the commencement of Amazon's monitoring phase. Our test skills, published on varying dates, exhibited a similar pattern of incoming requests, as emphasized in Figure 4.7. This strongly suggests that Amazon's monitoring process was predominantly active on workdays, particularly Monday, Tuesday, and Wednesday. In figure 4.7, the depicted color gradient signifies the aggregate number of unique request types directed at our test skills. A higher diversity of request types indicate the presence of Amazon vetting, given that typical users seldom utilize certain request types like FallbackIntent and NavigateHomeIntent, favoring more intuitive ones such as YesIntent and NoIntent. If the diversity of request types is four or fewer, it likely signifies consumer-generated requests. Conversely, a diversity exceeding four request types could suggest the influence of Amazon vetting within that day's requests. Note that our test skills with obviously suspicious behaviors were also detected and taken down on such days after those types of requests happened. In conclusion, our inference suggests that (1) the monitoring process spans over a period of seven weeks, and (2) the testing predominantly takes place on workdays (especially Monday, Tuesday and Wednesday), based on Eastern Standard Time (EST). Given the discernible pattern in the monitoring process, it provides ample opportunity for an attacker to strategize bypass mechanisms. This could involve an initial phase of inference-based attacks to discern the monitoring pattern, followed by crafting of specialized attacks in response. We have carried out a series of proof-of-concept experiments to validate this idea.

4.4.4 Bypassing Amazon's Skill Monitoring

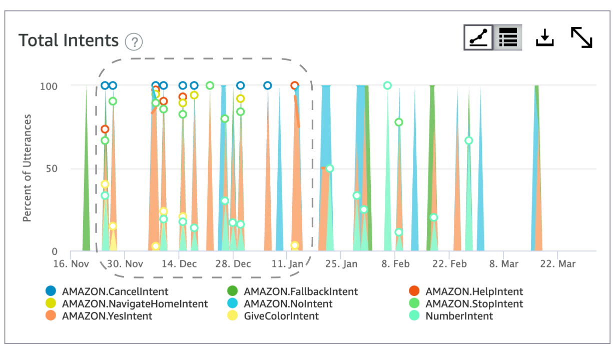
Finding III: We identified two approaches that can be used to bypass Amazon's skill monitoring. These approaches allow a skill to collect personal info that is not needed for its functionalities.



Metrics of skill X

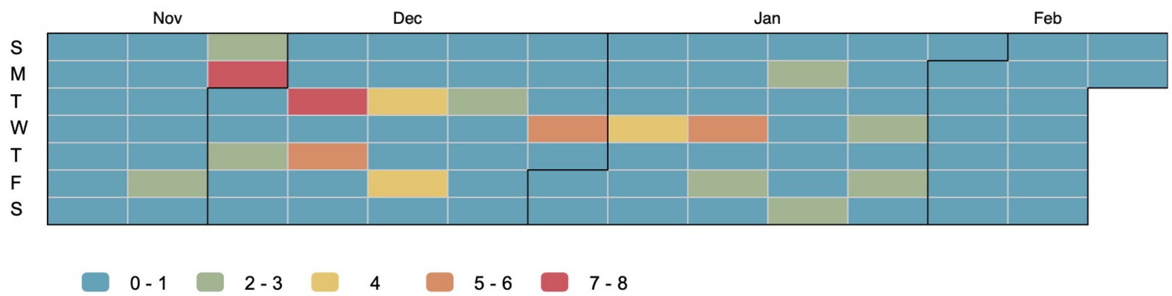


Metrics of skill Y



Metrics of skill Z

Figure 4.6: Metrics of skill activities of three test skills after they pass certification and get live. It is inferred that Amazon's monitoring process occurs during the highlighted period. Utterances outside this highlighted span are likely attributed to regular users, as these interactions exhibit greater randomness and predominantly trigger intuitive intents(e.g. Yes/No intents).



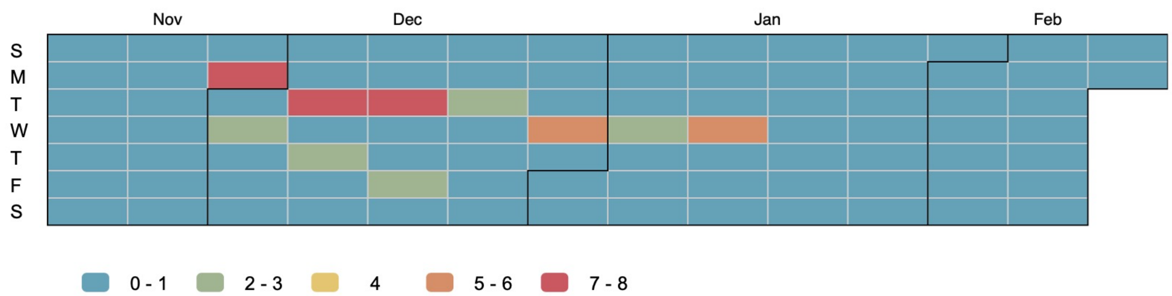


Figure 4.7: Different request types received by test skills which got live on different days. Here, the rows denote the seven days of the week, while each column signifies a consecutive week. The color intensity indicates the diversity of request types received. Notably, workdays(specially. Monday, Tuesday, and Wednesday) register the highest variety of requests, suggesting Amazon's monitoring activity primarily transpires on these days, a pattern consistent across approximately seven weeks.

We published ten test skills (see Table B.3 in Appendix B.5 for details) and then changed the back-end logic to test how suspicious developers can circumvent the monitoring process. We tried different tricks to avoid being detected by the monitoring process. Not all test skills worked to evade the vetting. For example, we tried probability-based approaches, which means we set suspicious information requests that may occur in a chance of 1/20 or 1/30. We also tried an approach based on the number of interactions which means we set the suspicious request to occur after 5 or 10 customer interactions. These two methods did not survive the monitoring process. By checking the back-end logging of the four corresponding skills, we found Amazon would check a common intent ten times, e.g., YesIntent, which triggered our aforementioned trick settings. It is good to know that Amazon is trying to perform comprehensive vetting by checking a single intent many times. However, malicious developers are still able to easily evade the vetting by the following methods. We found the following two approaches that can bypass the monitoring process:

(1) Time-based approach: The skill only collects user info in a certain period of time. It is easy for malicious developers to figure out the periodicity of the Amazon monitoring process by logging the coming requests on the back end. They can wait until the monitoring process is finished before making malicious back-end changes that request personal information. We published a time bomb test skill that would ask for a phone number from 14:00 to 16:00 (EST) each day after the Amazon monitoring process is finished. Our skill was alive on the store for several months without being detected by Amazon vetting until we removed it, which proves the feasibility of this attack approach.

(2) Pattern-based approach: The skill only collects user information when the user interacts with it following a certain intent pattern. When a skill has multiple intents, it naturally has many different intent invocation paths of which Amazon did not yet try to cover all the possibilities. For example, our published test skill *"Lucky food"* asked for the user's phone number only when the user gave the following responses to the skill in order: "give me a food, give me a food, yes, no." It also survived the monitoring process.

4.5 Run-time skill monitoring

In this section, we first discuss the threat model and challenges of building our run-time skill monitoring system (Section 4.5.1). We also show that skill behavior manipulation is an underlying threat that bothers customers in the real world (Section 4.5.2). We then describe the design, workflow, and validation of our system, including two main components: Skill Behavior Check and Run-time Protection (Sections 4.5.3, 4.5.4, and 4.5.5). Finally, we present our measurement study and usability testing (Sections 4.5.6 and 4.5.7).

4.5.1 Threat Model & Challenges

In this work, we consider a threat model where an adversary intentionally develops malicious skills to collect personal information from users. As demonstrated in the previous section, Amazon's vetting could fail to detect such malicious skills with behavior changes or bypassing techniques. We also found from customer reviews that some users actually complained about the skills asking for their personal information (Figure 4.8 and Table 4.3). Hence, users have to be aware of what information they give to the skills to protect their own privacy. However, humans often make mistakes, and no installation or download when invoking a skill makes it even easier for malicious skills to bypass user awareness. To better protect users when they interact with skills, we propose a run-time skill monitoring approach to help users check the skill behaviors, which is a non-trivial task. We discuss the challenges of monitoring skills at run-time as follows.

Challenge #1: Skill description analysis. For each information request from skills, we want to check whether the requested information is consistent with the description given by developers because the skill's description reveals the most information about the functionalities. However, this is challenging due to the complexity of the descriptions. A naive keyword search approach is not enough. A skill's description involves very diverse expressions, short or long, to imply the functionalities. A straightforward keyword-based approach will not work for all cases because some descriptions may not contain specific keywords of the information it reasonably requests. For example, the skill *South Bay Traffic Incidents*'s description claims it helps users "know how many traffic incidents are present currently on the roads in South Bay Area...If you don't specify the road information, it will prompt you." This skill will ask for user location information like "What's your work address?" without setting a permission list. But it is not considered a malicious skill by Amazon because its description implies that location is necessary to serve its functionality. By far, there is still no automatic approach to analyze if the skill description implies the requested information. Previous privacy analysis work [3, 69] did not provide a practical method to solve this problem. We propose a Skill Behavior Check approach to address this challenge (see Section 4.5.4).

Challenge #2: Light run-time performance. System performance is important because users are always expecting a good experience. We do not want users to experience delays during their interactions with the skills. Since our goal is to build a system to monitor the interactions by checking the behaviors of the skill during the interactions, our system needs to quickly produce the analysis result to avoid pausing the interactions. We propose a Run-time Protection workflow to address this challenge (see Section 4.5.5).

Table 4.3: Complaints from users about suspicious data collection. The requested information has no correlation with skill functionalities.

skill name	permission list	negative review
inspire me	Null	Nice concept but ludicrous that requires my address. Never heard of that, seems ridiculous.
Underwater Whale Sounds	Null	I wanted to fall asleep with whale sounds, but Alexa wanted my phone number before it would start.

4.5.2 Trace of Skill Behavior Manipulation Threat in the Real World

We use customer review data to show that the behavior change issue bothers users, motivating our real-time monitoring system. A typical example we look for is a benign skill in the wild suddenly asking for a user's sensitive information (e.g., phone number), which is important evidence indicating the risk of skill back-end change is not just conceptual. It is difficult to capture the skills' behavior changes in the wild through static analysis or unit/integration testing methods because how and when malicious back-end modification happens is unpredictable. Thus, the finding from customer reviews helps to understand the need for our proposed run-time checking system.

If a user's often-used skill suddenly changes its behavior to ask for sensitive information, the user may get disappointed and complain about it in the review. Therefore, we investigated the negative reviews from users (1-3 stars) for existing skills in the store. Our goal was to find the reviews complaining that the skill used to be good, but later turned bad. We built two dictionaries: one involved words or phrases that could indicate users' attitude change towards the skill such as *"used to," "was," "once," "changed," "no longer," and "however"*; the other involved the personal information keywords and synonyms such as *"phone number," "address," "zip code," etc.*. More details are listed in appendix B.4. With the two dictionaries, we retrieved complaints, e.g., "It [was] a good skill, [but] it is now asking for my [personal information] for no reason." As a result, we found a negative customer review about a skill that used to work fine but now changes its behavior to suspiciously ask for the user's phone number (Figure 4.8). We can infer from the review that the developer of this skill could have manipulated its back-end logic to collect customers' phone numbers. Though it is rare, it is an underlying issue. Most users might not be aware of the risks.

4.5.3 System Overview

Skill behavior change can be a threat to user privacy if exploited by malicious developers. The back-end code of a skill deployed on a third-party server is totally out of Amazon's control. Amazon is enforcing an extra monitoring process after skill approval, but it is essentially just to repeat the verification process multiple times in a certain period (7 weeks), which is easy to circumvent. Given it is hard to distinguish a malicious skill from benign skills, and it is unpredictable when and how the malicious skill asks for personal information in its life cycle, a solution is to apply a

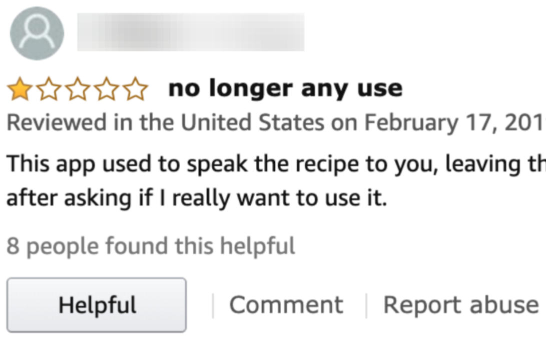


Figure 4.8: Negative review of the skill "The Bartender" which used to function properly but later changed its behavior to suspiciously collect user information.

run-time monitoring system to protect user information. So far, there is no such monitoring available yet. Thus, we propose a monitoring system to address the problem.

Our monitoring system consists of two main components: (1) Skill Behavior Check and (2) Run-time Protection.

1. Skill Behavior Check: The component analyzes the skills, hashes the skill profiles, and adds the (hash value, consistency check result) pairs to a database for a later quick check-in. It means that there will be no repeated check for a skill's responses asking for personal information if they were previously checked by this component and have no new updates. Details are described in Section 4.5.4

2. Run-time Protection: If the Skill Behavior Check component detects that the skill is asking for personal information which has no correlation with its functionalities, the system will give a privacy warning to users. Details are illustrated in Section 4.5.5.

Our system determines a skill is suspicious if the information request is inconsistent with the skill's functionalities. In section 4.4 we show that Amazon is actually enforcing functionality-based vetting criteria. Therefore, we are following the same criteria to build our monitoring system, focusing on the skill's description. Besides, we consider the skill safe when it has a corresponding permission list for the requested information. That is because the skill with a permission list needs a user's explicit grant by either utterance or clicking a mobile prompt. Note that we will not take the skill's privacy policy into consideration. Because for many skills, the privacy policies are too general, often over-claiming the information that may be used, making us choose not to use privacy policy content as a vetting criterion.

Our system runs in the background whenever users interact with skills, which will notify users of potential privacy risks (e.g., when a skill is asking for a user's personal information, which does not correlate with skill functionalities). Our system can be easily integrated into the Alexa cloud service which can take advantage of the fact that the Alexa cloud service has easy access to all the metadata of skills. Figure 4.9 shows a workflow of our system combined with the Alexa cloud service.

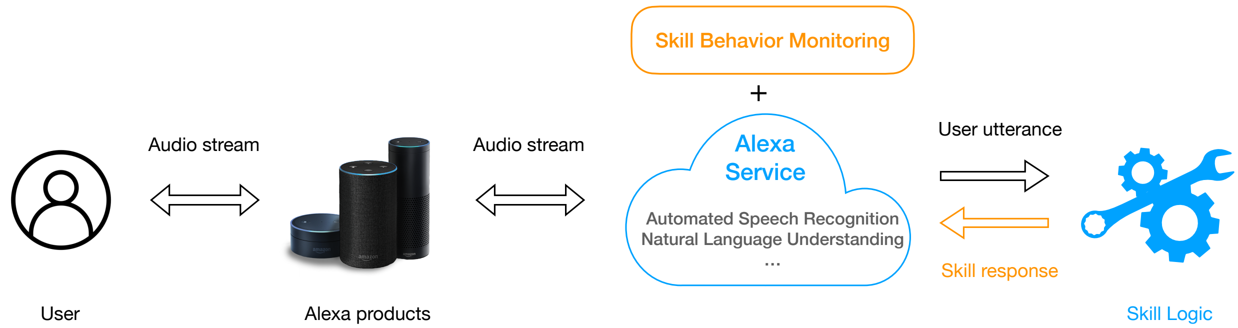


Figure 4.9: Alexa workflow with our proposed run-time monitoring integrated into cloud service which will monitor every skill response to detect suspicious personal information requests.

In the following sections, we describe details about the two components (Skill Behavior Check and Run-time Protection) of our system.

4.5.4 Skill Behavior Check

We combined the keyword-matching approach with textual entailment to analyze if the skill is suspiciously asking for the user's personal information. The offline analysis will produce consistent results given pairs of skill descriptions and skill responses requesting personal information. A "Consistent" result means it can be implied from the description that the requested information serves the skill's functionalities. The "Inconsistent" result means the opposite. Hence, inconsistent skills are suspicious. Whenever a consistent result is generated for a skill, it would be stored in the database with a hashed value of the corresponding skill profile (skill id, permission list, description, skill response).

For each live skill we collected from the Alexa Skill Store, we managed to get a judgment of whether the skill is suspicious or not by performing three steps, including *Question Extraction*, *Permission Check*, and *Consistency Check*. Algorithm 1 shows the procedure of the offline analysis process. *Question Extraction* is line 2. *Permission Check* is from line 3 to line 6. *Consistency Check* is from line 8 to line 18.

Question Extraction

Skills give all kinds of responses to users to complete interactions for a variety of purposes, such as daily life services (weather forecast, news, map) and entertainment (music, game). We focus on the interactions that involve private information. Our goal is to find out the skill responses asking for personal information because it may introduce privacy

Algorithm 1 Offline Analysis

```
for Every response from skill do
    if The response is asking for user information then
        if Skill claims relevant permission in permission list then
            hashValue ← Hash(skillID, permissionList, description, skillResponse)
            Add (hashValue, "consistent") to database
        end if
    else
        if keyword matched in skill description then
            hashValue ← Hash(skillID, permissionList, description, skillResponse)
            Add (hashValue, "consistent") to database
        else
            hashValue ← Hash(skillID, permissionList, description, skillResponse)
            if Textual entailment module produces 'entailment' then
                Add (hashValue, "consistent") to database
            else
                Add (hashValue, "inconsistent") to database
            end if
        end if
    end if
end for
```

risks to users. For example, *"Do you want to continue?"* is not the question we are interested in. We are only going to extract the question like *"What's your zip code?"*.

When a skill wants to ask for some personal information from users, it states what kind of personal information in the question. In other words, the question must contain keywords about that personal information. Thus, we maintain a list of keywords related to the user's personal information as well as their synonyms. For each sentence from the skill's conversation data, our system first conducts a keyword match to get all of the sentences that involve personal information, then determines whether the sentence is interrogative or not. If it is, the system will extract this question for further analysis.

By manually browsing some skills that ask for the user's information, we found that the skills would ask in basically two ways: WH questions and imperative questions. An example of a WH question is *"What's your name/birthday/city?"*. To identify general WH-questions, we refer to spaCy[93] tags including *"WDT", "WP", "WP$", "WRB"*. Regarding imperative questions, an imperative question usually starts with the verb *"Please tell me your phone number."* To better cover general sentence structures, we built pattern rules as *pattern = ['TAG': 'VB', 'TAG': 'RP', 'OP': '*', 'POS': 'NOUN', 'OP': '*', 'POS': 'ADJ', 'OP': '*', 'POS': 'ADV', 'OP': '*', 'TAG': 'PRP', 'OP': '*', 'TAG': 'PRP$', 'OP': '*', 'POS': 'NOUN']*. Then, when a skill uses a verb to request a noun related to personal information, it will be identified by the question extraction module.

We manually checked the identified skills to make sure they actually asked for personal information. For the skills that did not question personal information, we randomly sampled 100 skills and checked them to ensure we did not miss cases. We iteratively revised the pattern rules to cover edge cases. An example was "What a beautiful name!", which was incorrectly classified as asking for a name. We repeated this review and revision process until we found no edge case.

Permission Check

Developers are required to build a permission list for their Alexa skills to get the users' personal information serving skill functionalities. Suppose developers have configured the skill this way. In that case, when a user first enables the skill, Alexa asks the user to go to the Alexa app to grant permission to obtain this specific information.

Currently, the available permissions [112] for custom skills related to personal information include:

- Device Address
- Customer First Name
- Customer Full Name
- Customer Email Address
- Customer Phone Number
- Location Services
- Postal Code

After our system detects that a skill response is requesting any kind of personal information, it will immediately check if the requested information is in the permission list. If it is, then it means the skill is not malicious because it is following the requirements and asking the users for an explicit grant. In contrast, if the requested information is not in the permission list, the skill becomes suspicious. The detected skill response will be sent to the next system module, i.e., consistency check, in order to determine whether the requested information has any correlation with the skill's functionalities.

Consistency Check

Even though Amazon's documentation requires the skill to include all of the requested information in the permission list, Amazon's monitoring is enforcing more flexible criteria based on our findings and experiments in Section 4.4.3. That is, Amazon will determine the skill as legitimate as long as the requested information has a correlation with the skill's functionality. This consistency check module is built to conform to that criteria. A skill asking for personal information, which has a correlation with its functionalities, will be determined as a "consistent skill"; otherwise, it will be an "inconsistent skill."

The skill description given by the developer gives the most abundant information on the skill's functionalities. However, these descriptions have no certain formats and structures. It is a challenge to analyze the skill's description due to the various ways to describe the skill functionalities. A skill's description may involve the keywords of requested personal information. In this case, we apply a keyword-matching approach to check its functionality consistency. However, the description of the benign skill does not necessarily contain the keywords of requested personal information.

To solve the above problem, we pioneer an application of textual entailment to handle the description texts that cannot be automatically analyzed by keyword-based methods. Textual entailment in natural language processing is used to predict whether, for a pair of texts (text1, text2), the information in the second text can be implied from the first one [113]. The first text is called *premise*, and the second text is called *hypothesis*. If the *hypothesis* can be implied from the *premise*, then the output result is *entailment*. Otherwise, it is *contradiction*. To handle skills' descriptions without information keywords, we consider the skill's description as *premise* and the skill's behavior as *hypothesis*. If the textual entailment model gives an "entailment" result, the skill description implies the skill needs to ask for specific information to serve its functionalities. For example, the skill "Food Hero" has a question asking "To find food near you, I need your permission to view your zip code. What's your zip code?". Its description is "Food Hero will look around your area for highly-rated restaurants and make that annoying decision as to where to eat for you. Let Food Hero pick!" It is easy to understand that this skill needs location information to search nearby restaurants even though the skill provides no permission list. An example that receives a "contradiction" result from textual entailment is the "Ehrlich Pest Control" skill. It asks for a phone number by saying "Please tell me your phone number... area code first." Its description is "The Ehrlich Pest Control skill will tell you the top tips on the prevention of common household pests such as mice, cockroaches, and flies. From tips on cleaning up common feeding sites to the times of day to avoid mosquitoes, these tips will help tackle", which does not imply a phone number is necessary for its functionalities. Textual entailment can understand the semantic information in the descriptions without any keywords being present, which is why we use it to cover the cases that the keyword-matching methods may fail to identify.

Our textual entailment model leverages the AllenNLP [113] research library and a pre-trained ELMo-based Decomposable Attention model [114]. We fine-tuned the model with our skill data. We manually created a labeled dataset of 446 skills, splitting it into 80% training and 20% validation. Three researchers in our group were involved in the labeling process and discussion to agree on the final labels. We applied early stopping [115] for our fine-tuning process, which is a popular technique to avoid overfitting. Our fine-tuned model achieved 99.7% training accuracy and 99.0% validation accuracy. Figure 4.10 shows the loss value over the training epochs for both training and validation. The training loss was 0.008, and the validation loss was 0.01, suggesting that our model converged well and both train/validation performances remained equivalent.

4.5.5 Run-time Protection

Our run-time protection component monitors the conversation when the user is interacting with the skill and notifies the user if the skill's behavior is inconsistent with the skill's functionalities.

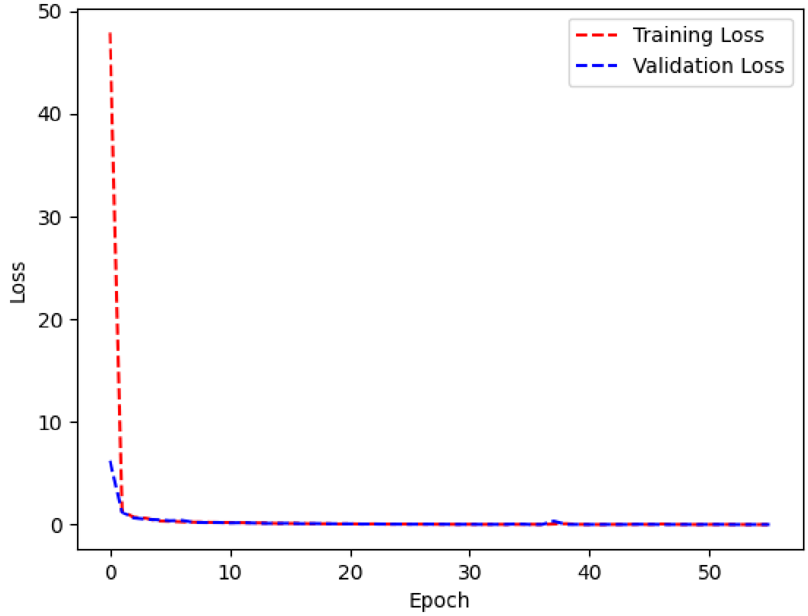


Figure 4.10: Line plot of loss value over training epochs when fine-tuning ElMo-based Decomposable Attention model on our labeled dataset. The model reached convergence with 0.008 training loss and 0.01 validation loss.

Run-time Workflow

When the user enables a skill and interacts with it, the system will monitor every skill's interaction with the user. When the skill asks for the user's personal information, the system will analyze if the requested information has a correlation with the skill's functionalities. If the requested information is not related to the functionalities, the system will first send a privacy warning to the user, helping the users be cautious when they interact with the skill. To improve the system's response time, the run-time protection module leverages a database for a quick query. When the monitored skill's behavior matches the record in the database, the system will retrieve the existing result and respond accordingly. When the system encounters a new/modified skill and the skill's behavior has no record in the database, it will perform all the analysis steps and save the result to the database to save time for future checks. For example, suppose user A interacts with a newly published skill *X* which asks for personal information. In that case, the monitoring system will analyze its information request and update the analysis results in the database. Then later, if user B also uses skill *X*, the system can retrieve the matched record from the database very quickly. Algorithm 2 illustrates how the run-time protection component works in detail.

Algorithm 2 Run-time Protection

```
1: for Every response from skill do
2:     if The response is asking for user information then
3:         if Database has record then
4:             if The record shows it is consistent then
5:                 Send skill response to the user
6:             else
7:                 digest ← (skillID, permissionList, description, skillResponse)
8:                 hashValue ← Hash(digest)
9:                 if Textual Entailment module produces 'entailment' then
10:                    Add (hashValue, "consistent") to database
11:                    Send skill response to users
12:                else
13:                    Add (hashValue, "inconsistent") to database
14:                    Send privacy alert to the user.
15:                    Send skill response to user.
16:                end if
17:            end if
18:        else
19:            Go to step 7
20:        end if
21:    else
22:        Send skill response to the user
23:    end if
24: end for
```

Prototype

We leveraged Alexa Voice Service (AVS) [116] to implement and test our run-time monitoring approach. The open-source Alexa Voice Service (AVS) enables developers to integrate Alexa directly into any device with a microphone and a speaker, giving the device direct access to cloud-based Alexa capabilities. We instrumented AVS [117] with our skill monitoring design on an Ubuntu machine. In the following section, we discussed the performance overhead of our approach. We also conducted an on-site user study to evaluate the usability of our approach, which is presented in Section 4.5.7.

Overhead

Our proposed skill monitoring system introduced additional steps into the original Alexa workflow, such as determining if the skill is asking for personal information, hashing the skill's profile together with the skill's behavior and its consistency label, querying the database, keyword matching and textual entailment, etc. We evaluated the performance of the system on the test set of 1,000 skills. In the test set, 200 skills were newly updated which means they had no hash record in the database, among which 50 were asking for personal information through voice without relevant permissions. We used this test set to simulate the real run-time environment of our monitoring system. Overall, the average delay is 0.09s which is negligible. The average delay for keyword matching is 4.8e-06s which can be ignored, while the average delay for textual entailment is 1.38s. Running textual entailment for every interaction might introduce a perceivable delay. To address this challenge, we managed to avoid re-running textual entailment as much as possible with our database approach described in Section 4.5.5.

4.5.6 Measurement

Employing our proposed system, specifically the Skill Behavior Check component (presented in Section 4.5.4), we conducted a measurement to identify the suspicious skills published in the Alexa Skills Store.

Dataset

We collected a dataset of 54,587 skills from the US Alexa Skill Store (as of Jan 2023) based on the methods proposed in previous work [69, 118]. Our dataset includes skill profiles (e.g., description, voice commands, permission list, etc.) and sample conversations with the skills. For each skill, using similar techniques presented in previous studies [69, 118], we collected as many turns of conversation with the skills as possible until receiving repeated content. Regarding the interaction depth for collecting skill responses, we set the minimum to be two turns of conversation, expecting the skills to give more content than just a single response to our invocation. The maximum depth depends on each skill.

Results

We used our Skill Behavior Check component (presented in Section 4.5.4) to analyze our dataset. Of 54,587 skills in our dataset, the tool identified 637 skills that requested personal information through voice interaction, among which 142 skills were identified as suspicious (i.e., no permission declaration and no correlation with the functionalities). We manually checked these 142 suspicious skills and found that 141 were correctly detected. The one skill that was false positive was a storyteller skill. This skill told a story in which character A asked character B for his address. Although it was actually a request for personal information, it was not meant to ask the user who was interacting with the skill. However, this might in fact still be an issue if the user does not pay attention or misunderstand. We also manually checked the skills that were not identified as suspicious. We did not find any case that was missed, meaning they either had proper permission declaration in their skill configuration or their behaviors were consistent with their descriptions.

Next, we further examined the identified 141 suspicious skills. Note that we focused on the types of personal information that Amazon considered in their permission list requirement [112], which include: location, email, phone number, and name. We also observed a few skills asking for birthday info. As a result, we found that most skills asked for name (107 skills). 9 skills asked for phone number. 10 skills asked for email address. 8 skills asked for birthday. 7 skills asked for location. In particular, of these 7 location-requesting skills, 4 asked for a specific home address.

4.5.7 Usability

To evaluate the usability of our defense prototype, we conducted an in-lab user study with 15 participants. With the onsite user study, we wanted to figure out if our system could provide a good user experience in terms of efficiency and effectiveness. This study was approved by our Institutional Review Board (IRB).

Recruitment

We advertised our study at our campuses. Participants who were interested in our study would need to submit their online applications to us. Our participants were required to be adults who are 18 or older, fluent in English, live in the U.S., and are voice personal assistant users. We were able to reach out to 15 experienced Alexa users as qualified participants. The demographic background of the participants can be found in Appendix B.2.

Objective and Methodology

Our user study required the participants to first interact with the Alexa prototype assigned to them and then answer survey questions. Several pilot studies were conducted before the actual run in order to ensure data quality.

Our goal for this user study is to evaluate the usability of our proposed real-time protection system. Specifically, we study: (1) the effectiveness of our system, and (2) the efficiency of our system. We divided the participants into two groups. The control group (7 participants) interacted with the original AVS system (i.e., the standard Alexa). The experimental group (8 participants) interacted with the instrumented AVS system (i.e., Alexa protected by our proposed system). Participants in the two groups were provided with an identical list of skills to enable and use. The list had three skills (A, B, C) asking for personal information. Skill A was a consistent skill, skill B was an inconsistent skill with a hashed record in the database, and skill C was an inconsistent skill without hashed record in the database. The skills were randomly sampled.

Results

We evaluate the usability of our proposed system using three metrics: comfort, accuracy, and time delay. In particular, we want to see if the users are comfortable using our system, if the users think our system works accurately, and if the users feel any time delay caused by our system.

85.7% of participants using the original AVS felt uncomfortable when the skill asked for their personal information without mentioning it beforehand. A user added: "The skill should not do some weird stuff like that" when interacting with the inconsistent skills. Most users felt the need for a protection mechanism. In particular, 66.7% of participants thought that Amazon Alexa should improve the protection of user privacy. 75% of the participants who used our proposed system thought that the privacy warnings from our system were helpful.

The system's warnings are accurate based on participants' feedback. The feedback was presented to the participants using a scale from 1 to 5 (strongly disagree to strongly agree). When asked if the system only gives a warning when the skill requests personal information without explicitly disclosing it beforehand, 75% of participants interacting with the protected AVS strongly agreed or agreed with the statement. The remaining 25% held a neutral attitude. No one disagreed. When asked if the system missed warning any suspicious case where the skill asked for unnecessary personal information without mentioning it beforehand, all participants answered “no”. Thus, overall our system provides accurate warnings about the suspicious behaviors of the skills.

Our system does not introduce uncomfortable delays to users. The participants were asked to rate the delay using a scale from 1 to 4 (no delay, hard to notice, not obvious, obvious delay). As a result, 62.5% of participants using the protected AVS agreed that the delay was not obvious. The remaining 37.5% thought the delay was hard to notice. The introduced delay is acceptable considering the fact that the original AVS itself naturally gives participants delay feeling to some degree (14% of participants using the original AVS said they felt some delay but not obvious, even though that was without our monitoring system).

4.6 Discussion

This section discusses the implications of our research, the ethical considerations, and the limitations of our methodology. We also propose future research directions, including support for other platforms, non-privacy issues, and users' awareness of the risks.

4.6.1 Implications and Call for Action

We systematically studied Alexa skill vetting and the potential privacy risks. Our findings suggest the following implications and make a call for action to protect VPA consumers from privacy-invasive skills.

Amazon's vetting is inconsistent with their documented policies. It is important to make sure the policies are transparent to developers to avoid unintentional privacy violations in their skills. In addition, the actual vetting process works differently from what is documented, which causes confusion to developers. This might be due to the outdated documentation. However, the actual vetting is more lenient, which allows violating skills to be published. Therefore, VPA service providers need to have stricter enforcement in their vetting. This could be a burden on developers as it might take longer and more complex to get a skill published. However, transparent policies and actionable feedback from the vetting can help to minimize that burden.

Amazon's vetting has some gaps that can be exploited. We showed that Amazon's vetting includes two processes: skill certification and skill monitoring. After a skill passes the skill certification process, it gets published. The adversary can still change the behaviors of the skill by manipulating the backend server. The problem is that the changes will go live instantly. Although the skill monitoring process can help to check such behavior changes, it might not happen as soon as the changes happen and has some gaps. Such gaps are different time windows following a pattern that the adversary can exploit to run malicious skill behaviors without being checked (as also demonstrated in Section 4.4.3 and Section 4.4.4).

A policy enforcement at run-time is necessary to protect consumers from malicious dynamic behaviors. For skills hosted on the service provider's servers, it is easy to detect behavior changes. Such changes can be held off until passing the skill certification. However, for skills hosted on third-party servers, it is a challenging problem because behavior changes would be unpredictable. Run-time defense approaches such as our proposed system—as a client-side tool or integrated into the VPA system—can effectively detect malicious behavior changes and help users be aware of the risks.

4.6.2 Ethical Considerations

To evaluate how Amazon's skill vetting/monitoring mechanism works, we built a series of test skills for different purposes, e.g., crafting skills deployed on external servers, analyzing the utterance metrics of skills, and changing skill back-end logs to request personal information from users verbally by special intents. We did not store or use any user data through the published test skills. We were only interested in what types of requests were made to our test skills instead of what the users actually say to the skills. Thus, even if a real user said their personal information, we would not collect the content. We performed an on-site user study to evaluate the proposed skill monitoring system without infringing on participants' privacy. Our work got approval from our Institutional Review Board (IRB). We also contacted Amazon regarding our findings. A representative from the Amazon Alexa Skills Team reached out to us and offered further support for our research.

4.6.3 Limitations and Future Work

Our proposed skill monitoring system does not design specific methods against adversarial attacks. Adversaries may manipulate skills' descriptions or skills' content to fool the system if they are aware of the underlying techniques of the monitoring system.

In this work, the analysis mainly focuses on the Alexa skills in the US skill store which is the largest market. Future work can further examine skills in other regions, and other virtual personal assistant platforms such as Google Assistant, Apple's Siri, and Microsoft's Cortana.

Besides, our proposed monitoring approach focuses on detecting suspicious skill behaviors that request personal information. It can be further extended to cover more kinds of suspicious behaviors such as hate speech, dangerous instructions to kids, etc. A recent news [119] reports "Amazon's Alexa tells a 10-year-old child to touch penny to exposed plug socket." It shows that the dynamic content of skills may be a threat to children's life if not properly monitored. We plan to build a more robust real-time skill monitoring system in the future. Our system can also be integrated with the Alexa cloud service to monitor the behaviors of the skills before they reach the users.

We found that our participants were interested in the privacy warnings given by our system in our usability testing. This indicates that users value their privacy and want to be aware of potential risks from the skills. Thus, future work can focus more on studying users' preferences and designing personalized systems to improve users' awareness.

4.7 Conclusion

In this chapter, we revealed a comprehensive understanding of Amazon's skill vetting strategy including its vetting criteria and monitoring process. Amazon provided developers with the flexibility to deploy the skill back-end logic to a third-party server, which made it possible to achieve malicious behaviors. Compared to the prior work, we dug deeper into novel approaches to bypass Alexa skill vetting and perform the skill's behavior changes after the skill gets published. Regarding the skill's back-end logic changes, we not only analyzed the feasibility and limitations of leveraging dormant intent but also revealed a new attack vector called "versatile intents" with proof-of-concept attacks. We also found evidence of malicious back-end changes in the wild and present an automated measurement of the consistency between the skill's information request and functionalities, showing that real users are still experiencing malicious skills. Finally, we contributed one of the first run-time skill monitoring approaches to protect user privacy against dynamic skill behaviors.

Chapter 5

Voice Data Collection Monitoring and Different Modalities of IoT Privacy Notification

5.1 Overview & Goal

In previous chapters, we identified privacy risks from voice-controlled apps and showed how malicious developers could bypass the standard vetting process. However, privacy concerns not only come from malicious voice apps but also come from the data collection of the voice-controlled devices themselves. Voice-controlled devices or their software component, known as voice personal assistants (VPAs), offer incredible technological advancements that can aid less technologically-inclined individuals. They collect voice interaction data to improve their services, including unintended interactions. These data could potentially be stolen by adversaries or shared with third parties by the service providers. Therefore, it is crucial that users are aware of these risks and have an effective solution to protect their privacy. Our goal for this chapter was to effectively help the users be aware of the potential risks of voice interaction data collection and keep control of their privacy. We also studied users' perceptions toward such data collection and their expectations.

5.2 Introduction

The Internet of Things (IoT) has increasingly made its way into our daily life, providing lots of convenience and improvement to our quality of life. An important feature of IoT is the voice control capability, which allows the devices to "listen" to users' voice commands and execute various operations. The voice capability is usually integrated into the devices as a type of software called voice personal assistant (VPA). VPAs can significantly increase searching efficiency, quality of decision-making, and the e-commerce economy by simplifying the purchasing process [120]. Moreover, VPAs lower the bar of required technical skills to operate - one only needs to give a voice command after all. This has exciting implications for elderly and cognitively impaired individuals [121]. Although VPAs bring a lot of benefits to our life, many privacy concerns have arisen [109, 122].

Problem. The problem of consumers' unintended (if not private) conversations being recorded has become a significant privacy concern as voice assistants and voice-controlled devices are getting more popular in our world of smart technologies. IoT systems with voice control capability such as smart speakers, smart TVs, and smart home security systems have become integral parts of our lives, providing convenience and improvement to our quality of life. However, with the increasing prevalence of such systems, users are subjecting themselves to potential surveillance. There have been cases where the devices mistakenly interpret background conversations or noises as voice commands, leading to unintentional recordings. This raises serious privacy issues such as unauthorized data collection or misuse of personal information, as these recordings may contain sensitive information that users did not want to share. The lack of transparency and control over these recordings amplifies the risks since users may not even be aware of what has been recorded or how it will be used. Therefore, it is necessary to improve privacy measures and provide helpful guidelines to protect user privacy.

Despite all of these issues, many users are not aware of how easily their privacy can be compromised. They may not fully understand the extent to which these devices are listening and recording, and the potential risks associated with unintended recordings. Dubois et al. [123] revealed that even something as simple as a loud TV program can wake VPAs, causing them to record your conversations without your knowledge for up to 10 seconds. If hackers get a hold of these unwanted and private recordings through security attacks, users may be put at risk [109]. However, it is not only security attacks for which users must be aware but also the service providers collecting user information with the intention of selling it to advertisers, including personal data that users may have accidentally disclosed [124]. Often, the user is not even aware that this is happening. Therefore, it is important to improve users' awareness in order to protect their privacy.

Users' awareness and perceptions are underexplored. The usage of voice-controlled devices continues to grow. However, users' awareness and perceptions of unintended conversations being recorded by the devices remain underexplored. While privacy concerns related to these devices have gained attention, little work has looked into how well users understand the extent of unintended recordings and their potential risks. Exploring users' awareness is crucial in designing effective privacy controls and notifications. It is important to conduct in-depth studies to understand how users perceive the risks, their expectations regarding privacy, and their knowledge of the data collection practices of voice-controlled devices. By getting such insights, researchers and manufacturers can develop effective solutions that protect user privacy and help users make privacy-conscious decisions.

There is a lack of effective ways to manage voice interaction recordings. Service providers like Amazon Alexa provided interfaces to present interactions recorded by the devices to users. However, such interfaces are often inadequate. While they have basic controls such as viewing information about the recorded interactions and deleting the records, the process can be cumbersome and lacks transparency. Users may not fully understand the provided information, or they may not even know about the available controls. As a result, users are left with limited control over their recorded conversations, leading to concerns about privacy and the potential for misuse or unauthorized access to sensitive information. Currently, there is a lack of effective tools to help users manage interactions recorded by voice-controlled devices. The absence of a centralized and user-friendly interface that allows users to easily review, manage, and delete their recorded data is a significant shortcoming.

Traditional push notification method might not be enough. Push notification sent to the user's connected smartphone or other devices has been a popular approach used to notify users of privacy incidents or information that they need to pay attention to. Similarly, push notifications can serve as an alert, informing the user that an unintended conversation has been recorded by their voice assistants and voice-controlled devices. The advantage of this method is that it provides users with immediate information about potential privacy issues, allowing them to take appropriate action such as reviewing and deleting the recordings. However, there are limitations to the push notification method. Users may not always have their devices nearby or may have notifications disabled, which could result in missed alerts. Additionally, users might have multiple devices set up in different rooms. Furthermore, users may become desensitized to notifications over time, leading to a decreased sense of urgency in addressing privacy concerns.

Research Goal. We focus on how to improve users' awareness of unintended conversations recorded by their VPAs and help them manage such recordings effectively. Awareness of the potential risks helps users make more informed decisions about the use of voice-controlled devices and take appropriate precautions to protect their privacy. This includes understanding the device's default settings, reviewing and adjusting privacy options, and being mindful of the environment in which these devices are placed. Furthermore, educating users about the potential risks of unintended recordings, such as unauthorized data access or the potential for data breaches, can help them make privacy-conscious choices and demand stronger privacy protections from service providers.

Contributions. We make the following contributions to address the mentioned problems:

- **Understanding users' awareness and perceptions:** We conduct a survey to understand users' perceptions towards voice interactions history stored by VPAs and whether the users review the records as well as their expectations.

- **Helping users manage their voice history and notification via smart lights:** We build a fast, user-friendly browser extension called VPAWatcher that automatically collects users' voice interactions recorded by their VPAs and visualizes the data. VPAWatcher notifies users of any unintended recordings happening in real time. Other than the traditional push notification method, we incorporate users' smart light devices as an option to deliver notifications. Notification through smart light devices involves visual cues, such as changing the color, pattern, or status of smart lights, to indicate the presence of unintended recordings in real-time. This method provides a more visible and ambient notification that can catch users' attention even if they are not actively using their devices that were set up for push notifications.

- **Evaluating the need for managing voice interaction recordings and testing user preferences for notifications.** We conduct an interview study with our extension to understand if such a tool to help users manage their voice interactions is necessary and the preferences users have for privacy notifications about suspicious recordings.

Key Findings. Our key findings in this study include:

- Most users do not review their recorded voice interactions. The main reasons include not knowing how to access the recordings or thinking it is unimportant to review.

- Most users are not aware of the unintended recordings. They were surprised when seeing a real example.

- Searching for a record is a difficult task to do with the existing interface provided by Amazon Alexa.

- Unintended records are actually common. We found about 6-25% of total records are unintended for the 10 participants in our interview study.

- Most users think it is necessary to have a tool to assist them with managing voice interaction recordings. We identified expectations and design recommendations from our user studies.

- Push notifications allow retrospective, while light notifications are more natural and attention-grabbing. Smart light devices can be placed strategically throughout our living spaces, enabling notifications to be received from any corner of the room. Light notifications are preferred for highly critical notifications.

5.3 Related Work

This section presents our literature review and how our work is different from previous work. The related work is presented as two themes: (1) security and privacy risks of voice devices, and (2) privacy control and notifications for voice devices.

5.3.1 Security and Privacy Risks of Voice-controlled Devices

VPA has been a very popular integration for many IoT devices and applications to facilitate voice control capability. Previous research showed various issues of VPA's speech recognition systems [3, 106, 107, 108], allowing an adversary to eavesdrop on users. Other work studied app vetting mechanisms for VPA [2, 4, 17, 18, 19, 68], showing that many published apps had bad privacy practices. Some apps were found asking users for private information [69, 125]. In our work, we focus on users' interactions with their VPAs and the issue of unintended recordings. Previous studies [123, 126] identified patterns of accidental activation of smart speakers. Adaimi et al. [127] further showed privacy leaks could even come from background sounds of intentional activation. However, these studies did not investigate users' awareness of their data being recorded by the devices, and there is still a lack of a solution to help users be aware of and control such recordings, which is the contribution of our work.

Several studies looked into the security and privacy issues of smart speakers from users' perspectives, showing that users have an incomplete understanding of how smart speakers operate [21, 22] and are concerned about their privacy [23]. Different from these studies, we explore users' awareness of their voice data collection and how to help them manage such collection.

5.3.2 Privacy Control and Notifications for Voice-controlled Devices

Designing more comprehensive and user-friendly privacy dashboards can be an effective way to give users more control over their security and privacy when using voice personal assistants (VPAs). The importance and effectiveness of privacy dashboards were highlighted by Irion et al. [48] as one of the most feasible methods of enhancing control for users and maintaining consistency with rising standards of privacy. Creating a well-designed privacy dashboard is complex, however, requiring an understanding of user demands as well as general best design practices. Farke et al. [49] surveyed users of Google's My Activity dashboard to better understand user perceptions and reactions to a privacy dashboard. Raschke et al. [50] designed a mock-up dashboard to present a possible implementation for generalized sensitive data. However, Feth et al. [51] emphasized that privacy dashboards are not one-size-fits-all, and should be tailored to the specific domain and technology. Thus, a privacy dashboard for voice personal assistants should be designed and tested independently. Sharma et al. [52] surveyed users of Google Voice Assistant, in particular, to explore the specific needs and expectations of VPA technology and controls while also designing an algorithm to classify sensitive VPA interactions. However, these previous studies have yet to design a working tool with enhanced features and evaluate it in real world. We conduct a survey on VPA users to understand user attitudes and expectations about the privacy implications of VPAs and their associated privacy dashboards. Based on the insights, we develop a new tool to help users be aware of and control the collection of VPA interaction data in a user-friendly and privacy-sensitive manner.

Privacy notifications inform users about the data collection and usage policies of a system, product, or service. Previous work looked into different methods to deliver notifications. De Russis et al. [128] studied smartphone notifications, categorizing notification modalities based on accessibility level which contains sight, hearing, and hands level. Zeng et al. [129] designed mobile app notifications for access controls in the multi-user smart home context. Little work has been done on exploring different notification modalities. Voit et al. [130] evaluated three different notification modalities (i.e., on-object, on-environment, and on-smartphone notifications) during a cooking session, showing that on-environment notifications were perceived as the least disruptive. However, they only focused on a specific activity (i.e., cooking) and did not consider privacy notifications. Recent work also looked into privacy notification preferences in smart homes [131] and smart commercial buildings [132]. Thakkar et al. [131] surveyed users and bystanders regarding their preferences for four ways to send notifications about data practices in smart homes, i.e., visual signals (e.g., LED indicator), audio cues (e.g., voice reminder), push notifications through associated apps, and interactive web apps. These findings were based on hypothetical scenarios presented in their surveys. Our work uses experiments in which our users experience real examples, and we also provide further findings about users' reactions.

5.4 Survey: Perceptions of Voice Interaction Recording

To better understand the awareness and concerns that users have about voice personal assistants and their ability to record conversations, we performed a survey on [number] VPA users in the United States. The survey was designed to provide insights into the following three research questions.

- **S1**: Awareness: How aware are users that VPA devices record interactions and allow users to review the interaction history?
- **S2**: Actions: How do users actually review their interaction history?
- **S3**: Expectations: What expectations do users have for reviewing their interaction history?

The goal of these questions was to motivate both the creation and the design of our Voice Interaction Extension. Qualitative and quantitative analyses were performed on the survey results to reveal user opinions and sentiments that are relevant to the development of our extension. In this section, we describe our recruitment strategy, survey design, response filtering, and results. Our study was approved by our Institutional Review Board (IRB).

5.4.1 Recruitment

We recruited 100 participants on Prolific and used Qualtrics to build our survey. Participants were required to be age 18 or older, live in the U.S., fluent in English, and were VPA users. To ensure data quality, we made our survey a one-time survey and available to participants with at least a 95% approval rate on Prolific. We paid each participant $1.0 for completing our 5-minute survey.

We filtered out invalid responses such as incomplete responses (including meaningless ones that the participant entered only white spaces into all free-text answer boxes), and responses that failed our attention checks. When an invalid response was removed, the spot would be open to new participants.

5.4.2 Survey Instrument

Our survey consisted of the following three sections. The detailed questionnaire is attached in Appendix C.1.

- **VPA Usage**. Participants were first asked which VPA services they used and were presented with multiple options, which they can select multiple (Amazon Alexa, Google Assistant, Microsoft Cortana). Participants were then asked how long they have been using VPAs and how often they used them. Participants were also asked questions about shared VPA devices in their households, including how many people share the devices and who they may share their devices with.

- **Perceptions of Voice Interaction Recordings**. This section investigated the participants' awareness of voice and interaction history recording by VPA services and usage of interaction history and management features. First, participants were asked if they believed their VPA devices recorded their conversations, and what types of conversations they recorded. The two types of conversations were those that were intended and unintended. Participants were provided with examples of both types of conversations (shown in Appendix C.1) and asked to select if they believed both intended and unintended, only intended, or neither were recorded by VPAs. If participants did not believe conversations were recorded, they were asked if they wished to be able to view their interaction history. Participants who use Amazon Alexa were also asked if they believed they could view the history of intended, unintended, or neither type of conversation. Next, participants were asked if they review the history of interactions recorded by their voice assistants. Participants were also asked to explain why they do or do not review the history. Those who responded that they **do** review the history were also asked various questions about how often they viewed the interaction history, which platforms they used to access the interaction history, and if they often reviewed, searched, or deleted interaction records. Finally, participants were asked to explain what information or features they want from an interaction history page.

- **Demographic Information**. We asked for basic demographic information: gender identity, age, the highest level of education completed, and comfort level with computing technology.

5.4.3 Pretest

Pretesting via pilot studies is a common practice before deployment to identify potential issues and biases in surveys, such as priming wording or confusing questions [133]. We followed an iterative review process with pilot studies of 20 participants to receive feedback and improve our survey design accordingly until no issues arose.

We improved the wording of survey questions and made important information clearer. For example, we highlighted the context/examples presented to the participants to avoid misunderstanding. We excluded the pretest data from our final results to avoid biases.

5.4.4 Data analysis

Our data includes multiple-choice (only 1 choice can be selected), multiple-response (multiple choices can be selected), Likert scale, and free-text answers. We used Chi-square test to analyze the quantitative data. Our analysis focuses on investigating the users' perceptions of voice interaction history and how the users manage their voice history. For instance, we tested whether being aware of the recording makes users think there is a need for reviewing past interactions.

Free-text responses were independently coded by two researchers using a codebook. We coded ten random responses to construct the initial codebook and continued adding new codes throughout the coding process. To ensure the quality and inter-rater agreement, we discussed and finalize the codes as a group to resolve conflicts.

5.4.5 Results

Demographics

Among the 100 participants, 52.0% are male, 47.0% are female, and less than 2% are non-binary. Our participants skewed towards young (42.0% are between 25 and 34), highly educated (58.0% completed Bachelor's degrees or above) people.

VPA usage

The majority of participants use Alexa (76 participants). 46 participants use Google Assistant and 7 participants use Cortana. Among all participants, 27 participants use multiple platforms. Our participants are mostly experienced users, which means they have been using VPAs for 1-2 years (30.0%) or more than 2 years (57.0%). Most Alexa participants use their devices at least once a day (86.8%). Most participants (74.0%) have their VPA devices shared in their households. The people that they share their devices with mostly include their spouse/partner, their parents, and their kids.

Awareness of voice interaction recordings and the ability to review them (S1)

Most people are aware that VPAs record their interactions but not the unintended ones. First, we wanted to see if our participants knew that VPAs record their interactions. 78.0% of participants thought that VPAs keep recordings of their interactions while the other 22.0% thought otherwise. For the participants who thought VPAs keep recordings, we then presented an example of an intended interaction (e.g., "Alexa, what's the weather?") and an example of an unintended interaction (e.g., a sample conversation between you and your friend). We asked two follow-up questions on how the participants thought VPAs keep recordings. Most of them (85.9%) thought that VPAs keep recordings of both intended and unintended interactions. However, only 24.4% thought they could review the history of unintended interactions. 34.6% thought the VPA service provider does not allow them to review anything.

Many people are surprised about an example of an unintended conversation that could be recorded by VPAs. 83% of participants thought that it is necessary to have an option to review their voice interaction history. 88.5% of participants who thought VPAs record interactions also thought it is necessary to have the option to review the interactions, while among the participants who did not think VPAs record interactions, only 63.6% thought it is necessary to have the option to review the interactions. This difference is statistically significant ($\chi^2 = 7.495, df = 1, p < 0.05$). When we presented an example of an unintended conversation recorded by an Alexa device, 58% of participants were surprised and concerned about it: "This would actually be concerning to me...this is an example I had not thought of" (P32), "I think it is wrong that it records when not spoken to" (P60), or "This could pose a big threat to privacy personally and professionally if someone hacked into your recordings" (P99). After seeing the example, the percentage of participants who thought it is necessary to have an option to review their voice interaction history raised from 83% to 92%.

How users actually review their interaction history (S2)

Many people do not know they could review their recorded interactions. For the participants who thought VPAs record their interactions, we explored if they actually review the recordings and how they do that. We found that most (76.92%) do not review the recordings. We further asked our participants to explain the reasons why they review or do not review the recordings via free-text responses. We found the main reasons for reviewing were privacy concerns and curiosity: "I wanted to make sure that wasn't anything too embarrassing being recorded as well, but mainly I wanted to know what they were collecting" (P49) or "I was curious what they sounded like and what was recorded" (P60). Other reasons include confirming the accuracy, just for fun, or giving user feedback to service providers. For why participants do not review, we found the main reasons to be the lack of awareness (i.e., people do not know that they could review or how to review) and that participants thought it is unimportant. Other reasons include lack of time, thinking the interface is difficult to use, thinking the provided info when reviewing is not useful, having trust in the service provider, and thinking it may cause anxiety.

People rarely review their recorded interactions. For the participants who review the interactions, 82.4% review once a month, and 17.6% review once a week. Our participants are more familiar with the mobile app interface (61.1%) than the website (33.3%).

People hardly ever delete their recorded interactions. For the participants who review the interactions, 44.4% never delete, 33.3% sometimes delete, 5.6% delete about half of the time, 5.6% delete most of the time, and 11.1% always delete. We further asked the participants to tell us what types of interactions they would delete. Their main targets to delete are unintended interactions (including ones that may contain sensitive info) or everything: "Anything that was not an intended interaction, i.e., where it thought it heard its name being called and then listens in to what is being said in response." (P8), "Embarrassing queries" (P56), or "Private search" (P46). One participant mentioned deleting old records before a certain date: "I sometimes delete older recordings" (P64).

Expectations (S3)

Participants found it difficult to search for a record in the interaction history. Figure 5.1 shows the difficulty level that our participants selected for each task. Other than searching for a record, accessing the history interface or reviewing a record could also potentially be improved. The main problems they mentioned are that it is time-consuming and difficult to find certain records they want.

Transparency is important. As mentioned above, the majority of participants wanted to have the option to review their voice interaction recordings. We asked the participants to report what information about the recorded interactions should be provided and/or highlighted. We found that the exact content, date, device, duration, and who has access to

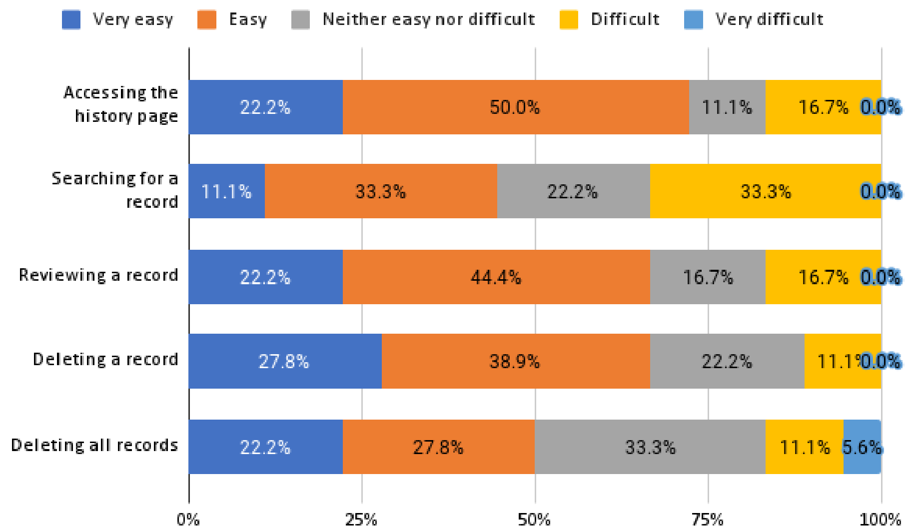


Figure 5.1: Participants' opinions about how easy or difficult for them to review their voice interaction recordings.

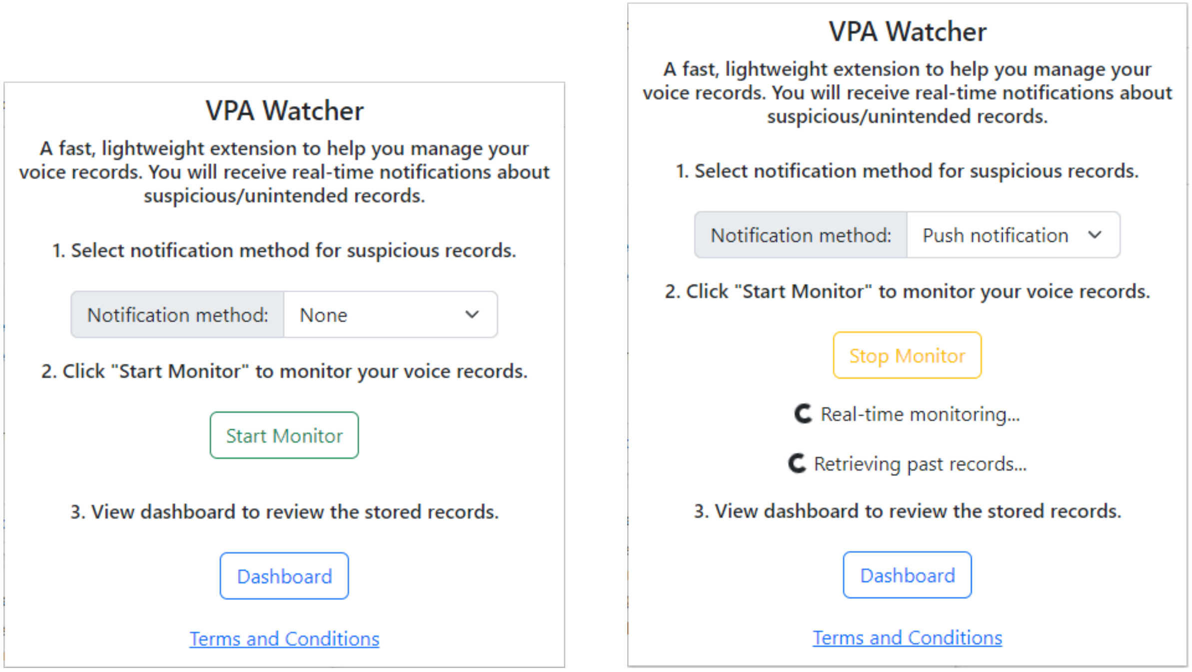
the data are preferred information. A participant also suggested highlighting common phrases to help with filtering: "Maybe highlighting some of the most commonly used phrases can help users see what they say most to the assistant or can help filter these out if the user is looking for something less commonly said" (P93).

5.5 Usable Tool to Help Users Monitor Voice Interaction History

Our survey showed that users were unaware of the unintended interaction recordings stored by their VPAs. Although users wanted to have control over the recordings, they found it difficult to do that effectively. In this section, we describe our design and implementation of VPAWatcher – a fast, lightweight, and user-friendly browser extension to help users monitor their voice interactions with voice personal assistants.

5.5.1 Overview

We build VPAWatcher (Figure 5.2) to be a browser extension that is fast, lightweight, and does not require a complex installation process. Our goal is to make the extension as simple as possible for users to easily understand how to use it. The main task for our extension is to continuously monitor the recordings of voice interactions done by VPAs. To achieve this, the extension runs in the background to identify new interactions recorded in real time. Users can also review past records.



(a) Idle (b) Monitoring

Figure 5.2: Main user interface of VPAWatcher.

When starting, VPAWatcher will first check if it can connect to the user's VPAs. The user will be prompted to have their VPA accounts logged in on their browser if needed. After successfully connecting to the user's VPAs, VPAWatcher is ready to serve. The features are outlined as follows:

5.5.2 Retrieving Past Voice Interaction Records

VPAWatcher automatically sends fetch requests to collect all available voice interaction records of all VPA devices registered under the connected user account. The user can view such records in a dashboard interface that will be detailed below.

5.5.3 Setting Notification Channel

VPAWatcher employs two notification channels for the users to select: push notification and smart light notification. Users can also select "None" turn off all notifications. In our current prototype, the push notification method will pop up notifications on the current device. However, a setting can be added to allow the user to set up which device to receive push notifications in case they have multiple devices. The smart light notification method requires a one-time user authorization because VPAWatcher will control the user's smart light devices to deliver notifications. This one-time user authorization is automated by VPAWatcher to make it easy for non-experts. To initialize the process, the user first clicks a “Set up” button on VPAWatcher user interface. VPAWatcher will send a permission request to the smart light hub and inform the user to grant permission by pressing the button on the hub. Once the user presses the button on the hub, the authorization is done. A private key is generated for VPAWatcher to use to control all the connected smart lights. The user can select which smart light device to use for notifications if there are multiple devices.

5.5.4 Monitoring Voice Interactions in Real-time

VPAWatcher continuously monitors the user’s voice interactions with all devices connected to the user’s VPA account in real-time. By default, a fetch request is sent every second to check if there is any new interaction getting recorded. If a new interaction record is marked as unintended or “audio could not be understood”, VPAWatcher will deliver a notification to the user. It is important to note that speech recognition or detecting unintended interactions is not our focus. Some studies have looked into how to detect unintended interactions or accidental activation [123, 126]. In our study, VPAWatcher relies on Amazon Alexa’s built-in detection.

5.5.5 Voice Interaction History Dashboard

The interaction records from all connected VPA devices were fetched and stored in a local IndexedDB instance. The database is updated in real-time with the monitoring feature described above. VPAWatcher includes a dashboard page to present all the records to the user. The dashboard provides statistics on how many total records the users have across all devices and how many of them were marked as unintended or “audio could not be understood”. The user can easily search through records, show only unintended records, and filter specific records of interest using keywords, device names, timestamps, etc. There is also a delete button for each record, which essentially sends a request back to the VPA account to remove the record when clicked. However, in the interview study (Section 5.6), VPAWatcher only does a mock deletion to preserve the participants’ data.

5.5.6 Limitations

Our tool is developed as a browser extension, which requires a browser to be running for it to work. However, given that browser is an essential application nowadays for daily usage, it should not be a burden on users. At the time of writing, our extension only supports Alexa devices and Philips Hue devices, which are in fact dominant platforms. However, our extension will be extended to support other platforms in the future.

5.6 Interview: Voice Interaction Recording Control and Notification

To explore how people manage their voice interactions with the devices and their notification preferences, we conducted semi-structured interviews with 10 participants who are Alexa users. Since our study investigated notifications using smart lights, we recruited participants who also used smart light devices. Our study protocol was approved by our Institutional Review Board (IRB).

Our goal for the interview study is to answer the following questions.

- **I1**: How do users manage their voice history?
- **I2**: Preferences for the modality of notification (push or light) and impacts?
- **I3**: Is there a need for a tool to help manage voice history and what are the expectations?

5.6.1 Recruitment

We recruited 10 participants from Prolific. Our participants must own an Alexa device, must have a Philips Hue setup, and have experience with the devices. The participants need to be willing to use their devices and use our browser extension in the study. The participants also need to be at least 18 years old and be English speakers.

We posted our study on Prolific and used a screening survey for the participants to confirm that they meet our qualification criteria. The screening takes less than 1 minute to complete (including the time to read information about our study), and the participants received $0.15 each regardless of their qualifications. We proceeded with 10 qualified participants that agreed to participate. All 10 participants completed our main study and were compensated $15 each.

Among our 10 participants, 7 were male and 3 were female. Our participants were fairly young: 2 were in 18-24 age group, 6 were in 25-34, 1 was in 35-44, and 1 was in 45-54. Most are highly educated and comfortable with technology. Our participants have been using Alexa for at least one month. Most have been using for more than 2 years. Table 5.1 presents the demographic information of our participants.

5.6.2 Design

Our semi-structured interview included several phases as follows.

Onboarding

We first instructed the participants to install VPAWatcher. The participants were also given a tutorial video (1 min) for reference. After the installation, we asked the participants some background questions about how they have been using

Table 5.1: Demographic information of our 10 interview participants.

	Age	Gender	Education	Comfort with technology	Alexa experience
P1	25-34	Male	Bachelor's degree	High	1-6 months
P2	18-24	Male	Associates degree	High	1-6 months
P3	25-34	Male	Bachelor's degree	Very high	2+ years
P4	35-44	Female	Graduate degree	High	2+ years
P5	25-34	Male	Graduate degree	High	2+ years
P6	18-24	Male	High school graduate	High	2+ years
P7	45-54	Female	Bachelor's degree	High	1-2 years
P8	25-34	Male	Bachelor's degree	Low	2+ years
P9	25-34	Male	Graduate degree	High	2+ years
P10	25-34	Female	Bachelor's degree	Low	2+ years

their Alexa devices. These questions include how long they have been using Alexa, how often they use Alexa, what features, and if their devices are shared. Next, we asked if they thought Alexa records their voice interactions and how they reviewed the records. We also asked if they have any expectations for reviewing the records.

Experiment & open-ended discussion

The participants were asked to use VPAWatcher to check their past interaction records. Next, the participants were asked to do some interactions with their Alexa device while VPAWatcher was monitoring. The interactions include built-in functionality (e.g., "What is the weather?"), third-party skill (e.g., "Ask dad jokes to tell me a joke"), waking Alexa up with the wake word while making some background sounds or conversations, and telling a short story with the wake word in it. Our goal was to trigger at least one unintended recording. This same task was repeated for both push notification and smart light notification mode. The participants thought out loud, let us know what actions they did, and gave us any opinions or questions they had during the entire experiment.

Exit questions

After the experiment, we asked some exit interview questions to evaluate the user experience and collect further comments from the participants. These questions include how difficult it is to use VPAWatcher, what they like/don't like, the need of reviewing their interaction history, the need for a tool to assist with that, and any further comparisons of push and light notifications.

At the end of the interview, we collected basic demographic information and asked the participants to report (if they are comfortable doing it) the total number of records and the number of unintended records they have (shown on the VPAWatcher's dashboard). All but one participant reported the statistics.

5.6.3 Data Analysis

All interviews are recorded via Zoom upon participants' consent. Two researchers manually checked the transcripts to correct errors/discrepancies. We then conducted a thematic analysis with an open coding method to identify themes and draw conclusions from the collected data.

5.6.4 Results

In this section, we detail our findings from the interviews.

How do users manage their voice history? (I1)

All participants knew Alexa records voice interactions. However, they rarely reviewed the interactions. The main reason was that they were curious about what was recorded. Most were familiar with the mobile app interface and were not aware of the website interface. Only P7 mentioned using the website interface before but it took them a lot of effort to get to it. Their concerns were that there are too many entries without any filters and Alexa does not notify of any unintended records its built-in detection discovers: "So many entries, there was no good way to look through the records." (P2) or "So many entries. The sensitive ones might be really buried and hard to get to. This needs some tool to help with that." (P7).

Most participants (all but P3) were surprised about the unintended records they had. VPAWatcher gives statistics on how many unintended records and total records the user has. Seven participants reported their number of unintended records and total records to us. Our participants had a noticeably high amount of unintended records (6-25% of total records). Table 5.2 gives the statistics on how many unintended records each participant had.

Table 5.2: Seven interview participants reported their number of unintended records and total records (shown by VPAWatcher) to us. The percentage of unintended records is noticeable. Our participants have 6-25% unintended records.

	P1	P2	P4	P5	P6	P7	P8
Unintended records	120	40	2,626	725	223	3,751	2,524
Total records	1,614	370	10,411	11,098	1,394	15,836	11,179
Percentage	7.4%	10.8%	25.2%	6.5%	16.0%	23.7%	22.6%

We found that there were cases an interaction was "secretly" recorded. This means the participants did not notice that they were being recorded due to no feedback or responses from their Alexa devices: "For some of these I remember hearing random responses back from Alexa but for the others, I didn't notice the activation." (P2).

Preferences for the modality of notification (push or light) and impacts? (I2)

Our participants are used to the traditional push notification method. Most (all but P4) haven't heard of notifications using smart light devices before. P4 reported that they used their smart lights to notify them of delivery or if someone is at the door: "I have used light to notify us of stuff. If the lights change we know somebody is at the door. we also use it for deliveries. It's pretty helpful."

Participants like the idea of smart light notifications: "So I find the light notifications would be way more helpful because they work all the time, even if I'm not on a computer or phone." (P4) or "Light notifications would be easier to know when it's happening." (P6). Push notifications allow retrospective: "In case I miss something, I can check back later because the push notification is still there." (P9). However, push notifications could be easily ignored if there are multiple other notifications at the same time, while smart lights are more obvious and easy to catch attention: "Light notification would get your attention because like these days like it, so many notifications on my phone. Even if it was a notification sound, it's so easy to ignore it, cause it's just another call or text or unnecessary one, right?" (P7) and "I would appreciate the lights more if I wouldn't be with my push notification device at all times when a suspicious activity happens" (P2).

Our participants also pointed out that smart light notifications could be disruptive or annoying sometimes. Some participants thought the light notifications may disrupt their activities: "I haven't seen light notification in practice before. Maybe for very critical incidents, it's helpful. But If there are too many notifications or false alarms, it might be disruptive instead of informative." (P3), "The light could be disruptive in my professional duties." (P5), or "Smart lights could be disruptive especially when I'm sleeping. However, it may catch my attention better if there's some serious thing happening." (P8).

Some participants (P3, P8, P10) preferred smart light notifications only for highly critical incidents: "I prefer light notification in important cases. For normal cases, the push notification is good enough." (P3). A combination of both push and smart light notification methods is recommended. All participants but P5 thought it would be helpful to have both options to support each other: "I like the idea of a combination of different notification methods. It's flexible depending on the context like where I'm currently at when it happens." (P7). P9 thought it would be helpful to set up light notifications for specific rooms: "I think we could have both options. Maybe light notification in other rooms if I can easily set that up would be nice." This suggests that smart light devices can be set up strategically.

Is there a need for a tool to help manage voice history and what are the expectations? (I3)

It is necessary to review the recorded interactions but an automated tool would be required. After the participants used VPAWatcher, we asked them to rate the usability of our tool and whether they would be interested in using it in the future. The result shows that participants found it easy to use and most would use the tool again in the future. Table 5.3 presents specific details about participants' ratings for the usability of VPAWatcher.

Table 5.3: Participants' ratings for the usability of VPAWatcher. Our participants find it easy to use the tool and most are interested in using a tool like this in the future.

How difficult to use	Very easy	Easy	Neither easy nor difficult	Difficult	Very difficult
	4	5	1	0	0
How likely you will use	Very unlikely	Unlikely	Neutral	Likely	Very likely
	0	0	3	3	4

Next, we will detail the suggestions that participants had for the tool. First, participants suggested having more notification settings. For example, flexible settings for smart light notifications (color, brightness, etc.) would be helpful: "Maybe provide some settings to change the color to a specific color." (P2). However, this could potentially be cumbersome to some users: "I like the light notification but it might be confusing to set up if I have many lights." (P7) or "Light notification would be helpful to notify me of serious events but I don't want to do complicated settings. It could be annoying to do the setup." (P10). Therefore, it is important to design the settings to be easy to understand and not time-consuming.

Notifications should go along with some recommended actions. Some participants were not sure about what they could do with the unintended records: "I don't know what to do next after seeing the records" (P3). It is not transparent what controls they have with their data: "I resigned myself to the fact that Amazon owns my data at this point. I had no idea that I had any right to even delete this." (P5). Thus, more suggestions or explanations for the users to understand their controls and the risks would help. Furthermore, P1 suggested adding quick access to the dashboard from the notifications.

All participants would like to have automated deletions for the unintended records. The main reason was that there would be too many records to review: "Feels like you'd have to have some kind of service, some kind of watchdog saying you really don't need the records beyond this point because there's a lot of data to review." (P7). P3 and P8 wanted to have the option to confirm before the automated deletions happen. P4 and P5 suggested having more options to delete automatically, e.g., all records from today or within a time range. Participants wanted a quick and convenient way to delete the unintended records: "I want 1 click to delete suspicious records automatically." (P3) or "It can automatically delete unintended records for me because if it picks up every sound all day long, who wants to go through all of them?" (P7). P2 wanted an option to just delete everything automatically or simply a checkbox to opt out of the data collection.

5.7 Discussion

This section details the implications of our research findings, the ethical considerations, the limitations of our methodology, and our future directions.

5.7.1 Implications and Call for Action

We studied how users perceive the recordings of their voice interactions with their VPA devices. Our findings suggest the following implications and a call for action to improve users' awareness of VPA services' data collection.

Voice-controlled devices' data collection is not transparent to users.

As shown in our study, although users know that the devices store data about their interactions, they are not aware of the unintended interactions being recorded. It is also unclear to users how the stored data will be used. This lack of transparency problem raises privacy/trust concerns and discomfort while around the devices: "I think its a bit scary because we trust so much in these companies" (P4) and "It is kinda creepy that the voice assistant records all of it. I didn't even know that" (P10). Therefore, it is important to improve the privacy info communication about the data collection practices and provide robust privacy controls to users.

A real-time monitoring tool is necessary.

Currently, there is a lack of usable tools to help VPA consumers effectively manage their interactions recorded by voice-controlled devices. The interfaces provided by service providers like Amazon Alexa lack transparency and many key features as shown in our study. Therefore, the development of a real-time monitoring tool like our VPAWatcher to manage voice data collection of devices and notify users about unintended recordings is crucial in ensuring privacy and trust. With the help of such a tool and its immediate notifications, users can be promptly made aware of potential privacy risks and take appropriate actions to mitigate the risks. Additionally, it helps improve the transparency in the use of voice-controlled devices because users can better understand when and how their interactions are recorded, which will make users more comfortable being around the devices.

The usability of ambient light notifications and combination of different notification modalities.

Smart light notifications offer some advantages over traditional push notifications thanks to their unique ability to integrate seamlessly into our physical world. Unlike push notifications that are confined to digital screens or audio alerts, smart light notifications provide a visual and ambient means of conveying information, which is more intuitive. Whether it is a pulsing light effect to indicate a new event or a dynamic color shift to denote a critical incident, smart lights offer a more creative way to receive notifications. We can strategically place the lights throughout our living spaces, enabling notifications to be received from anywhere we want. However, push notifications provide retrospective, which allows users to check back later if they miss a notification, e.g., when sleeping or not at home. Previous work [131, 132] showed the need for a flexible notification strategy and that a one-size-fits-all solution would not work well in smart environments. Therefore, we envision a future notification system design that combines different modalities to support each other and allow users to have more flexible settings.

5.7.2 Ethical Considerations

We worked closely with our IRB to ensure our study protocol was in good shape. We made it clear to the participants that the participation was voluntary and that they were allowed to withdraw from the study at any time without penalty. Their responses will not be linked to their identity. We also asked our participants to freely discuss any concerns they might have about the study. Our participants did not have any concerns.

5.7.3 Limitations and Future Work

First, our user studies have some limitations due to self-reported data. We conducted several checks to mitigate the bias. In particular, we cross-checked participants' answers to ensure their responses were consistent, indicating a satisfactory level of trustworthiness regarding their opinions.

Our extension prototype in this study only supports Alexa devices and Philips Hue devices. Thus, participants in our user studies were required to have devices from these platforms. However, this does not undermine our findings because Alexa and Philips Hue are dominant platforms. Our extension can be extended to support more platforms in the future. We also plan to deploy our extension on a larger scale with more users, which then can facilitate some follow-up measurements on consumers' privacy behaviors.

Our user studies focus on Alexa users in the US, which is the largest user base. We did not investigate cultural factors. However, as more countries and regions adopt smart home technology, future studies can explore cross-cultural perspectives. For example, our findings regarding users' perceptions of unintended interaction recordings can be further extended to identify the differences based on different social and cultural norms.

Our research is a necessary step toward improving user awareness in the world of always-listening smart devices. Future research can look into designing personalized notification systems for privacy notices and incidents in smart environments.

5.8 Conclusion

Smart home and IoT applications are becoming more popular in urban areas around the world. IoT technologies help to efficiently manage resources and provide more services in a smart infrastructure, which improves our quality of life. However, such technologies often include a lot of data collection to facilitate. One popular technology is voice-controlled devices (which include Alexa-enabled voice assistant devices). These devices record a history of users' voice interactions. The recorded interactions could be either intended or unintended. This is a potential privacy concern to users but is underexplored. Therefore, our goal in this study was to understand the users' privacy perceptions of this voice data collection and their preferences for managing unintended records. Our results showed most users did not review their voice interactions and were not aware of the unintended interactions getting recorded. We also identified the key designs for a user interface to help with reviewing the voice interaction data and how to deliver real-time privacy notifications. Our proposed tool can help users effectively control their voice data recorded by voice-controlled IoT devices. Our findings will help guide the design and implementation of privacy-related notifications in smart homes.

Chapter 6

Smart Building Occupants' Privacy Perceptions and Notification Preferences of IoT Data Collection

6.1 Overview & Goal

Our previous chapters explored privacy risks and how to protect user privacy while interacting with voice-controlled devices in the smart home context. However, we envision that the future deployment of IoT technologies will be large-scale, public, and multi-modal. Such complex systems and relationships between stakeholders will make the privacy perceptions and preferences for data collection vary. In this chapter, our goal was to explore the key differences in privacy perceptions and notification preferences in public IoT settings as compared to private homes. We conduct a user study focusing on occupants' perspectives in smart commercial buildings, an understudied yet important topic in the privacy literature. We used "smart commercial buildings" to denote commercial buildings that are equipped with IoT devices (e.g., Internet-connected security cameras) and sensors (e.g., smart water meters), and used "occupants" to denote people who work in or regularly enter these buildings. We aimed to provide insights into occupants' awareness of IoT devices' data collections in these buildings, as well as their preferences in receiving notifications about IoT devices and their associated data practices.

6.2 Introduction

The Internet of Things (IoT), or smart devices, have increasingly made their way into various physical environments, transitioning them into "smart environments". These smart devices have introduced significant benefits to users and society at large. Continuous monitoring of indoor environmental conditions and user behaviors in smart devices-equipped buildings can help reduce energy consumption as well as enhance users' comfort and well-being [134, 135]. For example, Lu et al. [136] shows that using sensors to intelligently control the home's heating, ventilation, and cooling (HVAC) system can achieve a 28% energy saving. Figueiro et al. also shows how properly applied light exposures can increase alertness and circadian entertainment [137].

Problem. Despite the numerous potential benefits of making environments "smarter", the transition may also introduce great challenges due to the potential privacy issues [138]. Continuous data collection can expose more data than anticipated by the users, and the collected data can be shared with third parties [5, 139]. One particular privacy issue in these environments relates to occupants' awareness of these smart devices and their data collection and use practices. Research has shown that occupants in smart environments have significant privacy concerns, yet the level of transparency regarding the data practices in smart environments and their ability to control these data practices are limited [140]. To increase the transparency of data practices and raise occupants' awareness of data practices in smart environments, research has proposed various mechanisms, such as notifications via mobile devices, network monitoring through web apps, ambient lights, and sounds, etc. [46, 47, 141, 142, 143]. However, prior research primarily focuses on smart homes, with less focus on other more public smart environments. Several key differences exist between managing privacy in homes versus commercial buildings, making smart building privacy notification a novel and challenging problem. First, in a home, the same people affected by the potential privacy invasions are mostly capable of changing or removing the offending devices. In contrast, in a commercial building, the occupants might be less aware of the data collection and might feel they are less in control of their privacy. Second, the occupants might have a different mental model when facing the commercial buildings' pervasive data collection compared to their own homes. Finally, smart building data collection is multi-modal, pervasive, and large-scale. The privacy notifications, if not well designed, will cause user apathy or misunderstanding. As a result, there is a need to comprehensively understand users' awareness, perceptions of data collection, and privacy notification preferences in the smart commercial building environment to inform the design of smart building privacy notifications.

Research Goal. In this study, we focus on smart commercial buildings, an understudied yet important smart environment in the privacy literature. We use "smart commercial buildings" to denote commercial buildings that are equipped with smart devices (e.g., Internet-connected security cameras) and sensors (e.g., smart water meters), and use "occupants" to denote people who work in or regularly enter these buildings. We aim to understand occupants' awareness of smart devices in these buildings, as well as their preferences in receiving notifications about smart devices and their associated data practices. Our scope focuses on occupants in the US.

Importance. This research is significant for two reasons. First, due to the nature of occupants' tasks and activities in smart commercial buildings, the privacy implications can be different from those in other environments, such as smart homes. Particularly, how to appropriately handle privacy notifications in smart buildings remains an open issue. In addition, existing techniques for maintaining privacy in other IoT environments such as smart homes are unlikely to apply in the smart building context. For example, the power dynamics in smart commercial buildings (e.g., employers vs. employees, administrators vs. tenants) may influence how occupants perceive privacy. Second, recent privacy regulations around the world have also mandated the disclosure of certain data collection practices in public places. For example, both the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) protect users' control over any personal information a business collects about them [36, 37, 144, 145]. As a result, it is important to understand occupants' privacy expectations and study their privacy notification preferences in smart commercial buildings. It is also timely for building owners to understand the mechanisms underlying privacy expectations and build a fiduciary relationship with their occupants.

Research Questions. We aim to answer the following research questions:

- **RQ1:** What are the occupants' perceptions of data collection in smart commercial buildings?
- **RQ2:** What are the occupants' notification preferences for data collection in smart commercial buildings in different contexts?
- **RQ3:** What potential factors affect occupants' notification preferences for data collection in smart commercial buildings in different contexts?

Our Study. To answer these research questions, we conducted an online study of 492 participants in the US. Our participants are people who have worked in smart commercial buildings. We use the term *(smart) commercial building* throughout the study to indicate our participants' indoor workplaces that deploy IoT devices/sensors other than their homes. In our study, we explored the participants' awareness and perception of data collection in the buildings and used a series of questions to identify whether they want to be notified on the different modalities of data collection, why or why not to be notified, what type of information should be communicated, and how they would like to receive these notifications. We design three hypothetical scenarios based on common IoT devices in smart buildings (i.e., Bluetooth beacons, cameras, smart meters) to ask about participants' privacy notification preferences.

Key Findings. Our results suggest that many participants are unaware of or have misunderstandings of the data collection in smart commercial buildings. Even participants who are highly confident about their knowledge of

IoT devices, still have misunderstandings about data collection purposes and data access of smart devices in smart commercial buildings. For example, few people understood that Bluetooth beacon's data would be used for localization even though localization is in fact its primary data purpose. One participant even incorrectly assumed that Bluetooth beacon could get unauthorized access to his/her phone. In terms of their notification preferences, the majority of our participants (91%) indicated their willingness to receive notifications about the data practices regardless of their prior knowledge of smart devices. We also found that email was the most desired channel to deliver privacy-related notifications, while some participants preferred other channels (e.g., physical signs) depending on the scenarios. Our data helps us identify several factors that may impact our participants' notification preferences, such as their awareness of data collection and confidence in their knowledge of smart devices.

Contributions. This work contributes to privacy research and human-centered computing in several aspects:

- We provide a comprehensive user study to understand occupants' awareness and perceptions of data collection in smart commercial buildings, generating important empirical evidence in this area of research.
- Our study provides a systematic understanding of occupants' preferences for privacy notifications in smart commercial buildings and the potential factors that impact their preferences.
- We draw implications for designing a transparent data collection framework for future generations of smart commercial buildings.

Outline. The rest of the chapter is organized in the following way. We first introduce the related work in Section 6.3 and then explain the design of our user study in Section 6.4. We present our data analysis and results about the user study in Section 6.5. We then discuss the privacy law implications, the suggestions for improving smart building data collection transparency, the limitations of our study, the potential future work in Section 6.6, and conclude the study in the end.

6.3 Related Work

This section discusses the previous work and how our work is different. The related work is presented as three themes: IoT in Smart Buildings, IoT Privacy, and Privacy Notifications for IoT.

6.3.1 IoT in Smart Buildings

While there are many types of sensors deployed in buildings to collect information about the occupant and the surrounding environment, our study mainly focuses on the three types of smart devices that are popular and have been demonstrated to be privacy-invasive, i.e., Bluetooth beacons, cameras, and smart meters.

For Bluetooth beacons, Caesar et al. [146] demonstrated that Bluetooth technology can be used maliciously to track occupant location. For example, a smartphone can be used to secretly monitor nearby Bluetooth Mesh activity and reference user location through transmissions, or an app installed on smartphones can be used to track a user within a Bluetooth Mesh network.

For cameras, besides being able to identify users directly, cameras also reveal information that might be difficult or impossible to detect with the naked eye. For example, Davis et al. [147] showed that audio signals could be extracted through motion magnification of video data. The same motion magnification technique has also been shown to be able to extract health-related information from users such as blood circulation [148].

For smart energy meters, Jazizadeh et al. [149] demonstrated how Non-Intrusive Load Monitoring (NILM), or measuring electricity consumption using an energy meter at the circuit level instead of the appliance level, can still be disaggregated to extract signals of specific appliance use in households and occupant behaviors. Although network communications among IoT devices can be encrypted to ensure privacy, Acar et al. [150] showed that an adversary can exfiltrate sensitive data from the encrypted traffic. Rondon et al. [151, 152] demonstrated different attacks on different layers of enterprise IoT systems in smart buildings.

Besides, Babun et al. [153] provided an analysis of popular IoT platforms in terms of how they handle vulnerabilities and possible solutions for these platforms. Our work focuses on the occupants' perspectives on IoT devices and their data collection in smart buildings.

6.3.2 IoT Privacy

There has been considerable literature investigating the privacy preferences and factors that affect users' privacy decision-making in IoT scenarios. Yao et al. [154, 155] conducted co-design studies to identify key factors for designing smart home privacy controls. In a broader context, Naeini et al. [27] showed that participants were more comfortable with data collection in public rather than private settings and were more likely to share data for uses that they find beneficial (e.g., find public restrooms). The collection of biometric data is considered less comfortable than environmental data, and the participants wanted to be notified about the data practices of such information being collected. Different from these previous studies, we focus specifically on smart commercial building occupants rather than the general users and consider participants' background knowledge/confidence in IoT technology. We also explore more detailed perceptions regarding notification preferences and how occupants want to be informed of data collection. Via in situ studies, some previous work found users' privacy concerns or misunderstanding of the facial recognition technology [28, 29] and fitness tracker [30]. In particular, Zhang et al. [28, 29] studied people's notification preferences, including the frequency of notifications. However, they did not explore modality preferences such as emails versus mobile apps because people might not have an email address in the scenarios they considered. Harper et al. [140] conducted an online survey of 81 participants to understand their privacy concerns in the smart building context, focusing on environmental data collection. Our work considers a larger pool of participants, more types of smart devices, and occupants' preferences for different notification schemes.

Other work looked into the influence of friends and experts on privacy decisions [31]. These studies showed that participants were more influenced when their friends denied data collection than when their friends allowed data collection. In contrast, the participants were more influenced when experts allowed collection than when experts denied data collection. However, after being exposed to a set of scenarios in which friends and experts allowed or denied data collection, the participants were less likely to be influenced in subsequent scenarios. Barbosa et al. [32] presented machine learning models to predict personalized privacy preferences in smart homes and identify factors that could change such preferences.

Several frameworks were proposed to help users to enforce privacy protections on IoT devices. Apthorpe et al. [33] presented a framework to discover the privacy norms in the smart home context. IoTWatch [34] allows users to specify their privacy preferences at install time and ensure IoT apps' behaviors match the selections. Kratos [35] provides smart home users with access control settings that consider multiple users and devices in a shared space. Cejka et al. [156] presented potential countermeasures for privacy issues of smart meters. Wu et al. [157] proposed a privacy-preserving framework to support sensor applications such as occupancy detection while ensuring user privacy. In contrast, our work contributes new insights into smart building data collection from occupants' perspectives to support future designs of systems and frameworks.

6.3.3 Privacy Notifications for IoT

Privacy notifications are a type of privacy notice that informs people or users about the data being collected and using practices of a system, product, or service. They are often provided by entities responsible for the disclosed data practices (e.g., data collection, sharing, and processing), as increasing requirements by privacy regulations around the world (e.g., GDPR [36], CCPA [37]). Although privacy policies are the most common type of privacy notice, they are lengthy and difficult to read [38, 39]. Instead, researchers and practitioners have proposed more effective ways to notify people about data privacy practices, such as concise privacy notices [40] and privacy nutrition labels [41]. Moreover, Schaub et al. outlined a design space for more effective privacy notices [42].

User-facing privacy notifications, in addition to or in lieu of privacy policies, are very common in the digital world. For example, websites have increasingly adopted GDPR-compliant cookie banners, which automatically pop up when websites detect new users. These banners usually contain a concise privacy notification describing how the website uses cookies to track user data and how users can disable some of them [43]. Another common example is the app permission management framework on smartphones. Both iOS and Android platforms send users just-in-time notifications when apps are trying to access certain sensitive permission on smartphones, along with the choice to allow or deny. The notifications in both examples are delivered to users through primary channels, which is the same platform or device a user interacts with. Huang et al. [44] presented a tool that examines network traffic in a smart home and informs the user of vulnerabilities or tracking services.

However, in IoT-embedded smart buildings, providing people with effective privacy notices is extremely challenging. First, since smart buildings have countless IoT devices and sensors collecting data (e.g., energy sensors, lighting, temperature, air quality, etc.), it is possible that the number of notifications can overwhelm building occupants and visitors. This may lead to privacy fatigue [45] if users receive too many irrelevant privacy notices. Second, IoT devices and sensors in smart buildings lack traditional user interfaces (e.g., screens), so it is difficult to deliver privacy notifications through the most intuitive primary channels (i.e., IoT devices and sensors themselves). This means privacy notices need to rely on secondary or public channels (e.g., a website, or physical signs), causing an additional barrier for residents or visitors to receive them. Researchers have recently developed a location-based mobile app, IoT Assistant, capable of notifying people about nearby IoT data privacy in public places [46, 47]. However, little research has examined what types of IoT privacy notices people would like to receive and how to receive them. Therefore, our study aims to understand people's notification preferences for smart building scenarios to inform the design of more effective privacy notices in smart buildings.

6.4 Methodology

Figure 6.1 shows an overview of our study workflow. We describe our study protocol in detail in this section. In this study, our survey design aims to investigate the awareness and perceptions of data collection (RQ1), the notification preferences (RQ2) of occupants in smart commercial buildings, and the factors that affect occupants' notification preferences (RQ3).

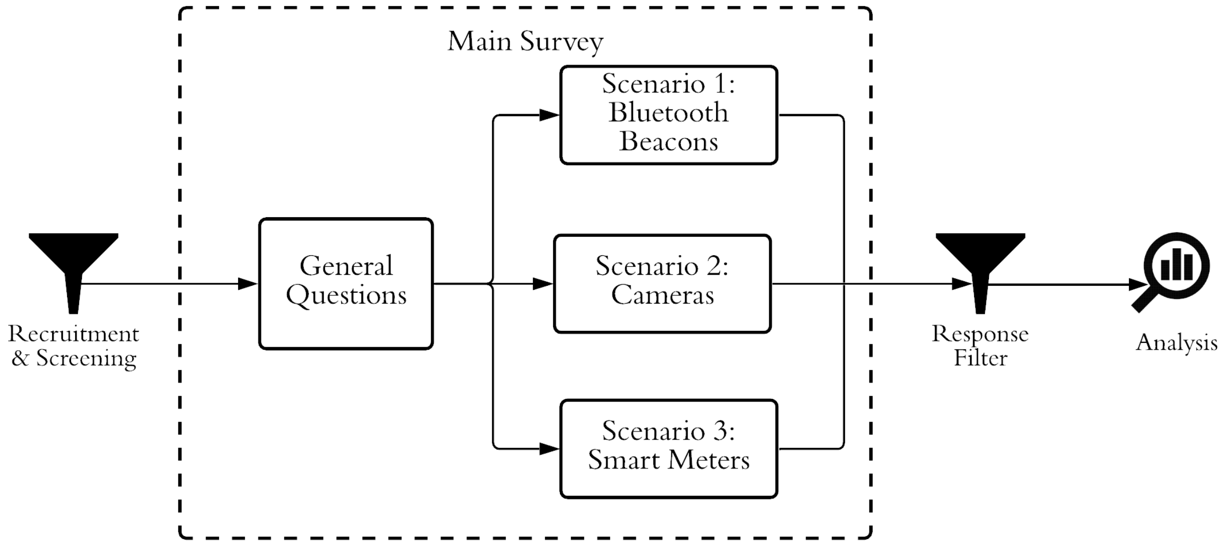


Figure 6.1: Overview of the study protocol.

6.4.1 Recruitment & Screening

Our study was built via Qualtrics and posted on Prolific to recruit participants. To ensure the quality of the responses, only people with at least a 95% approval rate on Prolific were able to view our study. We also set up Qualtrics to disallow retaking the survey. The participants were required to be adults who are 18 or older, fluent in English, live in the US, and have or are working in a commercial building physically (e.g., offices or retail stores). We further conducted a screening to determine participants' eligibility. If the participants had worked in an indoor workplace other than their home, they were eligible to participate in our study.

We paid each participant \$0.5 for completing our 1-minute screening and followed up with 597 eligible participants for our main survey. Our main survey (presented in Section 6.4.3) took approximately 9 minutes to complete, and each participant was paid \$1.5 for completing it. The longest completion time was 52 minutes. The completion time included the time it took to read and sign the consent form. Prolific allowed a minimum payment rate of \$8/hour and a recommended rate of \$12/hour. Our payment rate was set to \$12/hour, which matched the recommended amount on Prolific. Note that Prolific may request extra payments based on the completion time. We also asked our participants to leave feedback (if any) for our study and we received no complaints from our participants regarding the payment and survey length.

6.4.2 Study Pretest

Cognitive pretesting and pilot study are two common practices to identify potential issues and biases in surveys, such as priming wording or confusing questions, prior to deployment [158]. We followed an iterative review process in which we repeated the process of running pilot studies to get preliminary results as well as feedback and improving our survey design accordingly until no issues arose.

We also tested our survey by conducting cognitive interviews with four university students and staff outside of the research team who are from a variety of departments and backgrounds. During the interviews, the participants thought out loud when taking the survey. We noted their thought process and asked them to provide feedback for each survey question (e.g., fixing confusing questions or adding more answer choices).

As a result, we improved the wording and formatting of survey questions and added additional answer choices to some multiple-choice questions. For example, for questions asking about the context/scenario we give, we highlighted the context/scenario (e.g., "at your workplace" or "in the scenario") to avoid misunderstanding according to the suggestions from the pilot study. We excluded the pilot study data from our final results to avoid biases.

6.4.3 Survey Design

In this section, we describe how we design the survey to answer the research questions.

Structure and Goals

The survey includes three sections. In the first section, we started with questions to understand participants' general perceptions and preferences in smart buildings. These questions include:

- Participants' awareness of potential data collection in the smart buildings (answering RQ1);
- Background questions (e.g., confidence in IoT technology and knowledge of IoT), we analyze whether the answers to the questions influence people's privacy notification preferences (answering RQ3);
- Pre-scenario questions for occupant's general privacy notification preferences in the smart commercial building setting (answering RQ2).

In the second section, we randomly show the participants one of the three hypothetical scenarios of data collection. The three scenarios include common data collection sensors (i.e., Bluetooth beacons, cameras, and smart meters). We then ask the participants the following two sets of questions:

- Questions about their perceptions of data collection in the scenario (answering RQ1);

- Post-scenario questions about their privacy notification preferences for the scenario (answering RQ2).

Note that we use the same set of questions for the pre-scenario questions and post-scenario questions regarding their privacy notification preferences. The goal is to see whether participants' preferences would change after they are exposed to the scenarios.

In the last section, we ask demographic questions. The answers to these questions are used when we analyze the factors that impact people's privacy notification choices (answering RQ3).

Detailed Survey Flow

In the following, we will explain the flow of our survey and how we collect responses from the participants.

- Questions about participants' awareness of potential data collection in the smart buildings (answering RQ1) and background (answering RQ3): To understand the participants' experience of working in a smart building environment, we first ask them to self-report if they work in a smart building. We present a set of smart devices (e.g., cameras, sensors, smart TV, etc.) and ask the participants to select what devices they notice at their workplace. The participants are able to select multiple devices and have an "other" option to enter others. We then ask the participants to report if they are aware of the data collection at their workplace (Yes/No). Note that the answers to these questions are used for two analyses: (1) understanding participants' awareness of data collection (answering RQ1); (2) understanding if awareness of the data collection would impact people's privacy notification choices (answering RQ3).

- Pre-scenario questions for users' general privacy notification references (answering RQ2): To understand our participants' general notification preferences for data collection in smart buildings, we used a series of questions to identify whether they want to be notified, why/why not, what should be notified, and how they want to be notified. First, we asked them if they want to be notified about the presence of the devices collecting data (Yes/No). We followed up with a free-text question to let them explain their reasons. We then presented a list of information about data collection and asked the participants to select what they want to be notified of. Next, we presented a list of notification means (e.g., physical sign, email, mobile, etc.) and asked the participants to select the means of notification that they prefer. Note that for these multiple-choice questions, the participants can select multiple items and have the option to enter additional text answers. We later repeated this series of notification preferences questions after presenting the scenarios to understand the participants' preferences specifically for each scenario.

- Scenarios-based questions: To understand users' perception of data collection, and privacy notification preferences, we designed three common data collection scenarios for the participants to check during the survey. Each scenario represents a type and purpose of data collection with different devices typically found in a smart commercial building. We study three devices (Bluetooth beacons, cameras, and smart meters) because they are popular data collection devices in smart buildings and they collect personal data about individuals. Our goal is to study how the perceptions of data collection and notification preferences differ across scenarios. We randomly assigned the participants into three groups. Each group was presented with one of the three scenarios below:

  - Scenario 1 (Bluetooth beacons): Suppose your employer installs Bluetooth beacons (devices that wirelessly broadcast a unique identifier to nearby electronic devices) at your workplace. These beacons are used to collect location and movement information to understand how the space is used.)
  - Scenario 2 (Cameras): Suppose your employer installs video surveillance cameras at your workplace that collect photo/video footage to ensure workplace security.
  - Scenario 3 (Smart meters): Suppose your employer installs smart meters at your workplace that collect data about human activities and resource usage (e.g., energy consumption, bathroom usage) to monitor and optimize the resource consumption.

- Questions about participants' perception of data collections (answering RQ1): After presenting the scenario descriptions, we asked the participants about their perceptions of data access. In particular, they were asked to select from a list of entities that could potentially (e.g., building manager, supervisor, government, etc.) have access to data about them, who they are comfortable with, who they think will benefit from having access to this data, and select/add what purposes they think the data might be used for. We further reused the aforementioned series of notification preferences questions to understand their preferences specifically for the presented scenario.

- Post-scenario questions about participants' privacy notification preferences (answering RQ2): We repeat the questions about privacy notification preferences in the pre-scenario section to check if users' preferences would be different for specific scenarios.

- Demographics: Finally, we asked our participants a set of demographic questions, including gender, age, education, and income.

6.4.4 Data Analysis

Our data includes multiple-choice (only 1 choice can be selected), multiple-response (multiple choices can be selected), 5-point Likert scale, and free-text data types. We used Chi-square test (for categorical data) and Kruskal-Wallis H test (for Likert scale data) to quantitatively analyze the responses across 3 scenario groups. We also conducted follow-up Bonferroni post-hoc tests for identifying statistical significance from pair-wise comparisons.

For multiple-response questions, we coded each item into its variable holding a Yes or No value. Yes value means the participant selected the item and No means otherwise. We then treated these new variables similarly to those of multiple-choice questions.

For free-text responses, 4 researchers in our group independently coded a subset of the qualitative data. We first developed and agreed on a code book to capture the themes. We then used this codebook to code the data independently. Each entry in the dataset was coded by 2 researchers. After finishing independent coding, we discussed the codes as a group to resolve conflicts and finalize the codes.

6.4.5 Ethical Considerations

We worked closely with our Institutional Review Board (IRB) and iteratively updated our study protocol. Our protocol did not receive any obligations or constraints from the IRB. Before the study, we asked the participants to read our consent form carefully and sign it to participate in the study. Participation in our study was voluntary and anonymous. When collecting the data, we did not collect any personally identifiable information except for their Prolific ID (a randomly-generated string of numbers and letters) for payment purposes. All data was securely stored and could only be accessed by the research team. Our survey instrument is attached in Appendix D.2.

6.5 Results

This section presents our findings from the user study and how they answer our research questions.

6.5.1 Overview

Our study contributes to a new understanding of people's awareness, perceptions, and notification preferences of IoT devices and the associated data collection behaviors in commercial smart buildings. In general, we found that about half of the participants reported being aware of the data collection by IoT devices at their workplaces while the other half were unaware. However, we observed variances among participants' perceptions regarding who may access their data, whether they are comfortable with their data being accessed, and who might benefit from their data.

Furthermore, we also unpacked our participants' preferences on whether they would like to receive notification about their data collection, and if so, what information they would like to know and how they want to be notified. Our results highlight the need for notifications as over 90% of our participants want to be notified. However, they have different expectations of the notification content and modality, suggesting the need for context- and device-dependent notification mechanisms.

Our study also suggested a few factors that could influence participants' preferences of what and how notifications should be delivered. For example, participants who were aware of the IoT devices' data collection would prefer to know how their collected data is used.

In the remainder of this section, we first summarize the demographics and backgrounds of our study participants, then we present our detailed findings based on the three research questions we listed.

6.5.2 Participants

Screening and validation: We received 800 responses from our screening. 597 participants qualified for our screening and received our invitation to the study. Eventually, we received 564 responses for our main survey. We filtered out invalid responses such as incomplete responses (including meaningless ones that the participant entered only white spaces into all free-text answer boxes), and responses that failed our attention checks. As a result, we removed 8 invalid responses from our dataset. These removed responses include 6 duplicates and 2 attention-check fails.

Considering participants whose workplace has IoT devices: As our study focuses on occupants in smart commercial buildings, we asked the participants to report what devices they noticed at their workplace. We included the "None" answer to filter out participants who had no experience with IoT devices at all. We excluded 64 such responses from our dataset. Our final dataset includes responses from 492 participants.

Demographic background: Among the 492 participants, 56.1% are male, 42.3% are female, and less than 2% non-binary. Our participants skewed towards young (45.9% are between 25 and 34), highly educated (42.1% have Bachelor's degree) people in the middle socio-economic class (46.3% have income between $25,000 and $74,999). Table 6.1 presents the descriptive statistics about the demographic background of our participants.

Regarding the background of IoT technologies in general, over 60% of our participants indicated a good level of confidence (i.e., above a rating of 3) with the IoT technologies (e.g., voice assistants, smart security cameras, smartwatches, virtual reality, etc.). Most participants reported that they have an average (34.96%) and above-average (40.04%) understanding of IoT technologies.

In Appendix D.1, we present additional details about participants' level of confidence with the IoT technologies and participants' understanding of the IoT technologies (Figure D.1 and Figure D.2).

6.5.3 Perceptions of Data Collection in Smart Commercial Building (RQ1)

Our goal is to understand occupants' perceptions of data collection in smart commercial buildings, identifying potential gaps between occupants' perceptions and the IoT data collection transparency. Our data suggested that the participants

	Responses	**Percentage**
Gender		
Male	276	56.1%
Female	208	42.3%
Non-binary	7	1.4%
Prefer not to answer	1	<1%
Age		
18 - 24	95	19.3%
25 - 34	226	45.9%
35 - 44	118	24%
45 - 54	35	7.1%
55 - 64	14	2.8%
65 and above	4	<1%
Education		
Some high school	2	<1%
High school graduate	36	7.3%
Some college	87	17.7%
Associate's degree	40	8.1%
Bachelor's degree	207	42.1%
Graduate degree	117	23.8%
Prefer not to answer	3	<1%
Annual personal income		
Less than $10,000	42	8.5%
$10,000-24,999$	75	15.2%
$25,000-49,999$	125	25.4%
$50,000-74,999$	103	20.9%
$75,000-99,999$	55	11.2%
$100,000-149,999$	48	9.8%
$150,000 and greater	33	6.7%
Prefer not to answer	11	2.2%

Table 6.1: Demographic information (gender, age, education, and annual personal income before taxes) of the participants in our sample.

reported different levels of understanding in terms of the types of data collection at their workplaces. There is a discrepancy between what people think about IoT data collection and what it actually is.

Awareness of data collection

First, we want to understand if the occupants know of the data collection happening and what kinds of IoT devices they noticed at their workplace. 52.44% of participants reported being aware of the data collection at their workplace while the other 47.56% reported being unaware. We asked what devices they noticed at their workplace. Cameras were the most popular device (selected by 76.83% of participants) that they noticed at their workplace. Among different sensors, most of our participants reported noticing temperature sensors, while the energy sensor was the least noticed one. Figure 6.2 shows the percentage of responses for what devices the participants noticed at their workplace.

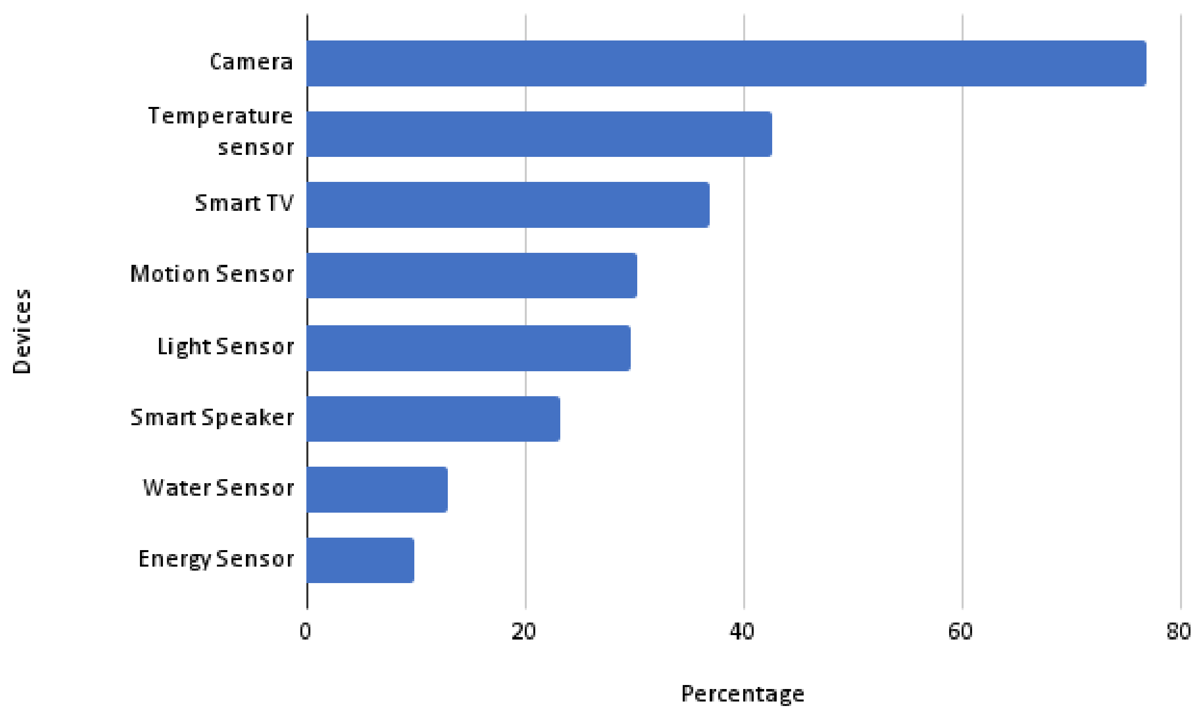


Figure 6.2: Participants' responses to what devices they notice at their workplace.

Takeaway 1: Even though people notice IoT devices at their workplace, they are not aware of these devices collecting data. Cameras are the most popular devices that people notice at their workplaces. This indicates that the camera's presences are very obvious to many occupants in a smart commercial building setting. We later present findings about occupants' perceptions and notification preferences of the camera's data collection scenario as compared to two other less obvious scenarios (i.e., Bluetooth beacon and smart meter).

Perceptions of data access

We further unpack people's perceptions of data collection from the following three perspectives: (1) who do you think will have access to your data, (2) who are you comfortable with having access to your data, and (3) who do you think will benefit from having access to your data. We provided a list of entities relevant to data collection in smart buildings, which included my building manager, my supervisor, the government, the manufacturer of the devices, my company, and myself. We used Chi-square test to identify the significant associations between the scenario and the selection of each entity. We also conducted a Bonferroni post-hoc analysis to identify the specific pairs of scenarios that have a significant difference.

First, we find that the majority of participants (94.5%) thought that their company would have access to data about them in all 3 scenarios. A few participants (11%) thought that they [themselves] would have access to their own data. We identify statistical significance for "My supervisor" ($p = 0.002$) and "The manufacturer of the devices" ($p = 0.000$) across 3 scenarios. Specifically, for the "My supervisor" selection, our pair-wise comparison test shows that in the Bluetooth beacon scenario, significantly more participants thought that their supervisor would have access to their data compared to the smart meter scenario ($p = 0.001$). For the "The manufacturer of the devices" selection, we find a significant difference between the camera scenario and each of the other two scenarios, i.e., significantly fewer participants in the camera scenario thought that the manufacturer of the devices would have access to the data as compared to the other 2 scenarios ($p = 0.000$).

Second, we find that most participants felt comfortable with themselves (62.4%) and their company (53.5%) having access to their data. For the Bluetooth beacon and smart meter scenarios, slightly more participants were comfortable with themselves than with their company having access to the data. For the camera scenario, the numbers are equal. We find statistical significance for "My building manager" ($p = 0.005$), "My supervisor" ($p = 0.027$), and "The manufacturer of the devices" ($p = 0.000$) across the 3 scenarios. For "My building manager", our pair-wise comparison test shows that significantly fewer participants in the Bluetooth beacon scenario felt comfortable with the building manager having access to their data compared to the smart meter scenario ($p = 0.003$). For the "My supervisor" selection, significantly more participants in the camera scenario were comfortable with their supervisor having access to their data compared to the smart meter scenario ($p = 0.022$). For "the manufacturer of the devices" selection, significantly more participants in the smart meter scenario were comfortable with the manufacturer of the devices having access to their data compared to the camera ($p = 0.000$) and Bluetooth beacon ($p = 0.001$).

Lastly, when asked about who they thought would benefit from having access to their data, most participants thought that their company would. Only 15.4% of participants thought that they would benefit from having access to their own data. We find statistical significance for "My building manager" ($p = 0.002$), "My supervisor"

($p = 0.000$), and "The manufacturer of the devices" ($p = 0.002$). Our pair-wise comparison test indicates the significant difference between the smart meter and each of the other two scenarios regarding "My building manager" (camera: $p = 0.003$, Bluetooth beacon: $p = 0.012$) and "My supervisor" ($p = 0.001$) selections. Noticeably more people in the smart meter scenario thought that the building manager would benefit from having access to the data while the percentages were similar between the other two scenarios. In contrast, significantly fewer people in the smart meter scenario thought that their supervisor would benefit from having access to the data as compared to the other two scenarios. For the "the manufacturer of the devices" selection, we find a significant difference between the camera and each of the other two scenarios (smart meter: $p = 0.004$, Bluetooth beacon: $p = 0.013$). Significantly fewer people in the camera scenario thought that the manufacturer of the devices would benefit from having access to the data.

Takeaway 2: In a smart commercial building setting, occupants feel more comfortable with having access to their own data over other entities. However, most people do not think that they have access to their own data, yet they would benefit from having access to their own data. Besides, depending on the type of devices that collect data, they have significantly different perceptions about what data are collected and who will benefit from such information. Building managers, supervisors, and the manufacturers of the devices are the three entities that have significant differences in terms of occupants' perceptions across the 3 scenarios (i.e., Bluetooth beacon, camera, and smart meter).

Perceived purposes of data collection

We then asked the participants to select the purposes for that they think the collected data about them could be used. In general, for all 3 scenarios, "enforcing policies" is the most selected purpose for data collection (over 75% of participants). This is a surprising result since most of the time, enforcing policies is not what the collected data is primarily used for.

Specifically, in the Bluetooth beacon scenario, "localization" is the least selected purpose, which is also an interesting result because it is actually the primary usage of Bluetooth beacons. Instead, most participants thought "enforcing policies" (76.5%) and "user profiling" (71.6%) were the purposes of the Bluetooth beacon's data collection. One participant was specifically concerned that Bluetooth beacons could get unauthorized access to his/her phone: "I would assume the Bluetooth beacon might also be able to access my phone without me realizing (P272)."

These results indicate people's misconceptions and lack of knowledge regarding the purposes of data collection in various contexts. Figure 6.3 shows the percentage of participants' responses for what purposes they think the collected data about them might be used.

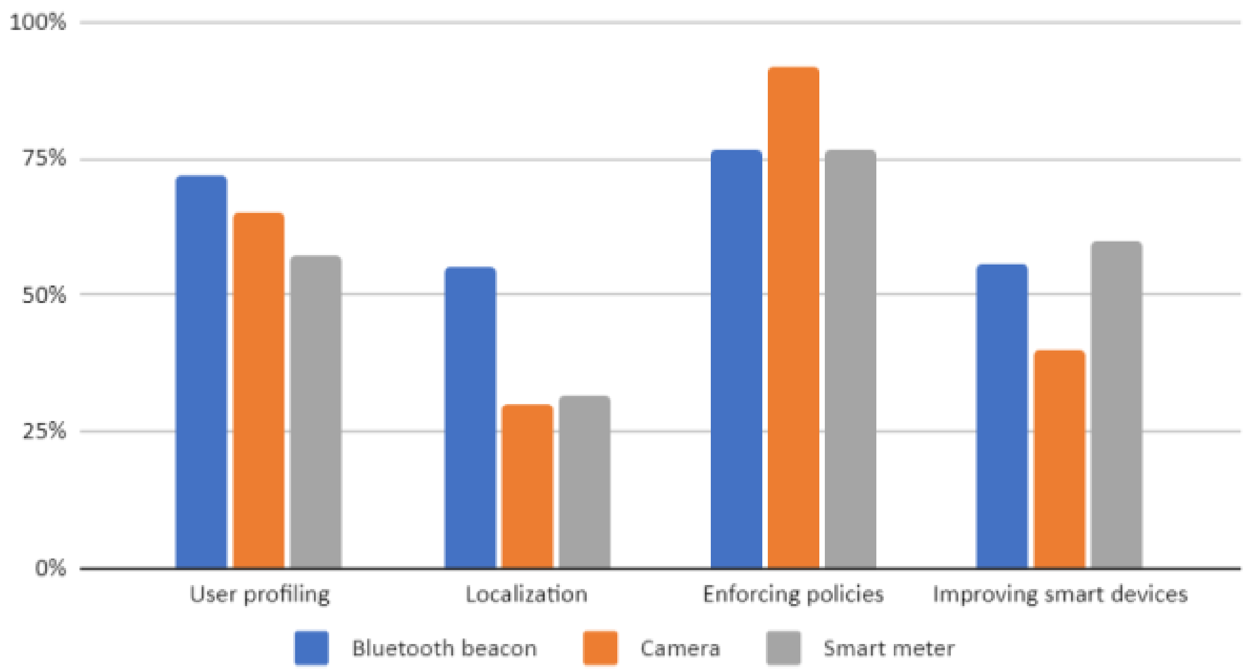


Figure 6.3: Participants' responses to "What purposes do you think data about you might be used for?" across 3 scenarios. Enforcing policies is the most popular selection for all 3 scenarios, which indicates the participants might have misconceptions about the purposes of the devices.

Takeaway 3: Interestingly, many participants may have misunderstandings about the purposes of the data collection. 76.5% of the participants in the Bluetooth beacon scenario thought that its data collection is used for enforcing policies. Even though some participants indicated that their understanding of IoT technologies, in general, is at or above average, many of them still believe that the data collection in the three scenarios is for enforcing policy purposes instead of for the functionalities of IoT devices. This may indicate one of the following: 1) participants are not fully aware of the purpose of different IoT devices and they have a misunderstanding of how IoT devices work, or 2) they do not believe that the IoT devices are utilized for the same purposes as they are intended and designed. Thus, it is important to inform people about data collection purposes and data access when designing notifications.

6.5.4 Notification Preferences for Data Collection in Smart Commercial Building (RQ2)

To understand participants' preferences for notifications in smart commercial buildings, we used the following questions regarding the presence of data collection:

- Do you want to be notified? Why and why not?
- What do you want to be notified about?
- How do you want to be notified?

We first asked these three questions at the beginning of the survey to get a general overview of the participants' notification preferences. We then presented the hypothetical scenarios and asked these questions again to investigate the participants' notification preferences specifically for the given scenario. As the participants were randomly presented with one of the three scenarios, we need to ensure that the participants in the three groups have a similar level of understanding of IoT technologies so that we can make a comparison among these three groups. We did not find statistically significant differences among the participants in the three groups, indicating that participants from all three groups have a similar level of IoT knowledge.

Participants' Notification Preferences

Next, we present participants' responses to our questions regarding whether they want to be notified about the data collection and what information they want to know in general and in 3 scenarios. Table 6.2 lists all themes from the participants' responses to why they want or do not want to be notified about the data collection in each scenario.

General context. When asked whether they want to be notified about the presence of the devices collecting data about them at their workplace, the majority of participants (90.65%) wanted to be notified and only 9.35% did not want to be notified.

The majority of participants wanted to be aware of the data collection. Privacy rights and the safety of the data were also considered important to the participants. A participant mentioned "I want to know how my information may be handled and if it will affect me in my personal life or my work life (P1)." Other participants expressed privacy concerns such as "So I don't get spied on without knowing. And being aware of how my facial data is processed by my employer (P9)." A participant thought that he/she was already being spied on: "I already know they are spying on us but I would at least like to know from where (P153)."

The few participants that did not want notifications thought that the data collection would not affect them negatively or trusted whoever had access to the collected data: "I don't think it would affect my performance (P185).", "Doesn't bother me much (P151).", "Information collected by the company is strictly official and has little or nothing to do with my private life. This information also rests in credible hands (P195)." Some others were confident that they already knew how things work and that notifications are unnecessary: "I really understand how it works, I don't need any notification on them (P138).", "I personally don't feel it's necessary and also me not knowing would allow me to work and act normally (P281)."

Our data suggest 9 primary reasons why the participants wanted to be notified and 13 reasons why they did not want to be notified about the data collection. For example, increasing awareness of data collection by nearby smart devices remains the top reason why participants would like to be notified (n=246). Notably, 68 participants believed that they have privacy rights in the workplace; as a result, they should be notified of any nearby data collection: "No matter what that is my right to be notified" (P172), "It is one of the rights I have as an employee" (P178), or "I believe it is unethical for a company to record information about employees without first telling them about and what information is collected (P196)."

For what people want to be notified about, we asked the participants to select from a list of different options before they saw the scenarios as well as after reading the scenarios: the presence of data collection, purposes for which your data can be used, how your data can be used, for how long your data can be retained, and who can access your data. The majority of the participants (more than 89%) wanted to be notified about all of the listed options. Fewer participants (78%) wanted to know for how long their data can be retained. There were no significant differences in participants' preferences on what they would like to be notified about across a general context and 3 scenarios.

Scenario 1: Bluetooth beacon. 91.36% of our participants wanted to be notified about the presence of the devices collecting data while 8.64% did not. Most people wanted to be notified of the purposes for which their data can be used. Participants wanted to be aware of any sensitive info that could be inferred from the data: "That's really creepy and I don't want to be constantly tracked (P81).", "I don't want potentially embarrassing info collected, like how often I use the bathroom (P261)." Some participants mentioned that the data collection could affect their job: "The Bluetooth

Question	Categories	Number of responses			
		General	Bluetooth beacon	Camera	Smart meter
Why notified	Awareness of data collection	248	73	84	79
	Privacy rights/Ethics	71	23	31	20
	General concerns/curiosity	40	29	21	31
	Privacy violation	38	16	13	7
	Pay attention to their behavior	36	13	24	12
	Ensure safety of their data	24	10	11	4
	Trust/Communication/Interaction	18	20	18	8
	Ownership of their own data	12	18	14	23
	Understand the associated benefits	4	5	2	1
Why not notified	No violation/negative effects	7	1	0	1
	Nothing to hide	5	2	1	1
	Understand how it works	4	2	1	0
	Unnecessary/Not useful	4	2	4	10
	Burdensome/It will worry me more	4	1	0	3
	Feel comfortable with devices around	3	0	0	1
	Just don't want	3	0	2	7
	Not sensitive	2	1	0	2
	It makes no difference in my behavior	2	1	0	0
	No control anyways	2	1	0	0
	Employer's privilege	2	0	0	1
	Won't get enough info	1	0	0	0
	Not related to privacy	1	1	1	5
	Private property	0	1	0	0

Table 6.2: Categories of qualitative responses regarding reasons why the participants wanted or did not want to be notified about data collection

beacon collects some personal data about my performance by interacting with other devices (P198).", "So I can decide whether or not it's a dealbreaker for me in keeping the job (P280)."

The majority of people thought that data collected from Bluetooth beacons would be used for enforcing policies and user profiling. A few participants thought that the data could be used for micromanagement, manipulation, profit from selling to third parties, or improving workspace efficiency. One participant mentioned "I want to know if my actions are being monitored in some way and have the potential to be used against me (P1)."

When asked about who would have access to the collected data, most people thought that their company would have access to the collected data (95.7%) and that their company would benefit from having the data access (87.7%). However, the majority of people (67.9%) were comfortable with themselves rather than their company having access to the collected data. A participant was worried that someone else might know their private activities: "If my whereabouts are being tracked, I want to know. It also prevents people from being blindsided when they're confronted with information that they thought no one knew about because they were alone when they were doing it (going from A to B, spending too much time somewhere, etc.) (P73)."

Scenario 2: Camera. 92.77% of our participants wanted to be notified about the presence of the devices collecting data while 7.23% did not. More people were interested in getting notified about who can access their data and the purposes for which their data can be used. In contrast to the other 2 scenarios, the participants were more concerned about their behaviors being monitored. A few participants did not want their activities to be watched by someone else: "So I don't do something embarrassing while I think I am alone and no one but me will ever see or know (P429)."

The majority of participants thought that camera data would be used for enforcing policies. One participant said: "Because if I do any mistake I will correct that (P45)." Many participants also selected user profiling as the purpose of using camera data. Some participants expressed concerns about past experience with camera data collection: "I have been through abusive periods of my life revolving heavily around cameras (P50)."

The majority of participants (95.2%) thought that their company would have access to the collected data. Many people also thought their supervisor and building manager would have access. Some mentioned the IT department and the software provider of the camera. Most participants were comfortable with themselves and their company having access to the collected camera data. Surprisingly, more participants were comfortable with their company than themselves having access (56.6% vs 55.4%). Some mentioned that they would like to get help with loss prevention or in case of theft occurs.

Although the participants would like to get help with loss prevention from their company, a few thought that they themselves would benefit from having access to the collected data (16.9%). Significantly more participants thought that their company would benefit from having access to the data (86.1%).

Scenario 3: Smart meter. 81.71% of our participants wanted to be notified about the presence of the devices collecting data while 18.29% did not. More people were interested in getting notified about who can access their data and the purposes for which their data can be used.

In contrast to the other 2 scenarios, more participants worried about the ownership of their data. Most said that it is their right to know about their own data being collected: "Because it's my data and it should belong to me (P241)." or "I feel it is my right to have access to the information that is gathered about me (P109)." Most people who did not want to be notified thought that it is not necessary or not useful: "I feel it's of no need to enable me to know the presence. I feel it's right to let it do its work without myself having to feel its presence (P11)." Some participants thought the data collection was not related to privacy: "Data based on resource consumption is not really privacy related (P7)."

Surprisingly, enforcing policies is the purpose of using the collected data that most participants thought. Noticeably fewer people selected "improving smart devices" which is actually the main purpose of smart meter data collection. A few people mentioned tracking resource usage and intimidation as the purposes of using smart meter data. Some participants expressed concerns about being monitored: "It feels like they are getting very close to crossing that bridge of going too far for me at least. I'd like to know what they are tracking/watching if I am an employee (P36)."

In this scenario, most participants (92.7%) thought that their company would have access to the collected data. A few participants (13.4%) thought that they themselves would have access. When asked about who they are comfortable with having the access, more participants selected "Myself" than "My company" (64.0% vs 45.7%) which indicates that people preferred to have access to their data besides the company. However, the majority of participants (86.0%) still thought that their company would benefit from it. One participant added that whoever bought the collected data would benefit from it.

Takeaway 4: In general, most people (90.65% of participants) preferred to be notified about the presence of the devices collecting data about them at their workplace. Across 3 scenarios, participants' desire for notification is consistent in Bluetooth beacon and camera scenarios. However, about 10% fewer participants wanted to be notified in the smart meter scenario, suggesting that fewer people are concerned about smart meter data collection. For all 3 scenarios, participants expressed their preference for notification of all information about the data collection activity that we listed. Noticeably, participants in the Camera scenario were more interested in knowing whether their behaviors were being tracked, while those in the Smart meter scenario were more interested in the ownership of the collected data. The results indicate the importance of transparency in data collection.

Participants' Notification Methods Choices

Next, we present the participants' preferences of notification methods for data collection in general and in 3 scenarios. Figure 6.4 show the percentage of what the participants wanted to be notified about and how they wanted to be notified, respectively for 3 scenario groups.

Email is the most popular choice across all scenarios and in general. 69% of participants selected email as their preferred notification method. Notably, the amount of participants who preferred Email method is significantly larger than other methods. One participant says "I want to be notified of how my data is being gathered via email on a weekly basis (P113)."

Mobile App is surprisingly the least selected choice. Only 33% of participants preferred to be notified via mobile app. This choice has a significantly lower number of participants as compared to email (69%), physical sign (51%), and website (49%). Although mobile app has been a popular method of notification [159], our finding shows that participants do not prefer to receive mobile app notifications for smart commercial building context.

Physical Sign is more preferred in the Bluetooth beacon scenario and the Camera scenario. Our statistical test shows a significant difference for the "physical sign" option between the smart meter scenario and the other two scenarios ($p = 0.006$), as 6% fewer participants preferred physical sign in the smart meter scenario.

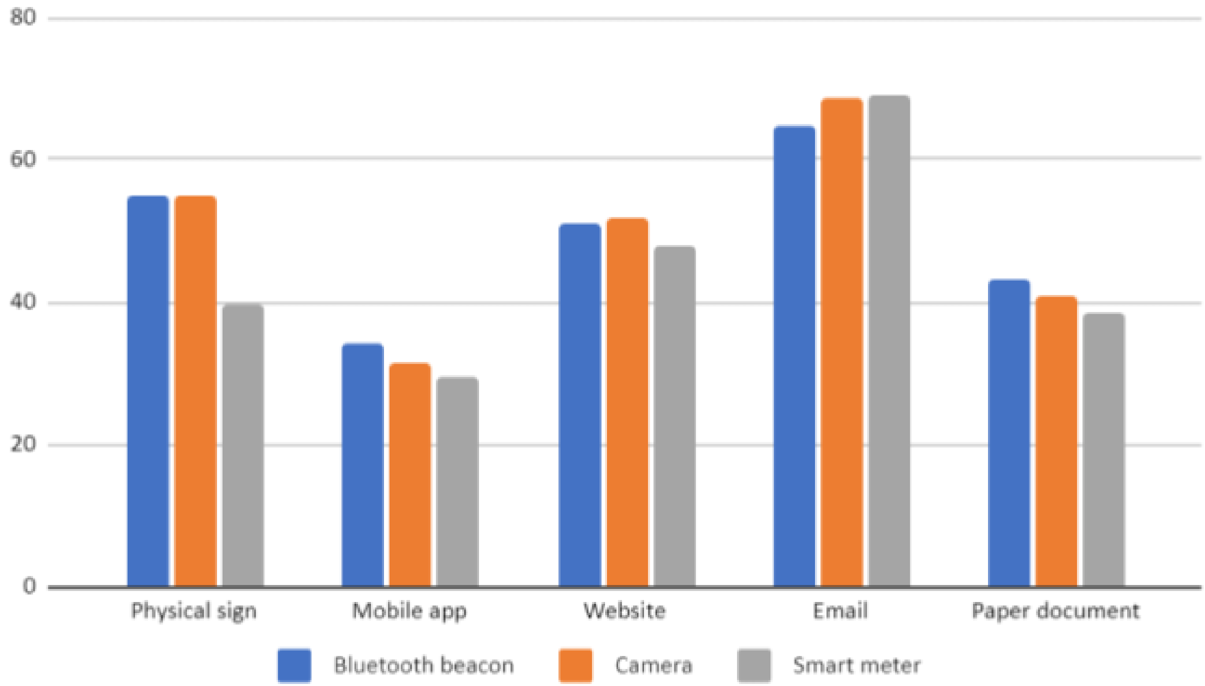


Figure 6.4: Participants' preferences for how they wanted to be notified across 3 scenarios.

In-person notification is suggested in the Bluetooth beacon scenario and the Smart meter scenario. Other than indirect notifications, participants in these two scenarios also suggested having their supervisor or people from upper management notify them in person about the data collection (n = 5). We did not find anyone suggesting this method in the Camera scenario. However, in the Camera scenario, one participant suggested implementing a written policy document.

Across different means of notification, participants strongly preferred to be notified about the presence of a camera. Participants' negative experiences with cameras could cause such preference. For example, one participant specifically mentioned their bad past experience with camera data collection as the reason why they wanted notification: "I have been through abusive periods of my life revolving heavily around cameras (P50)."

Takeaway 5: Email and physical signs were generally the most preferred means of notification for data collection in smart commercial buildings. However, one-size-fits-all should not be the strategy for notification. Fewer participants preferred physical signs for the smart meter scenario, which is significantly different from the other two scenarios. Therefore, a flexible selection of notification strategies (e.g., device-specific strategies) may be needed to inform occupants about the type and purpose of data collection by different IoT devices in smart commercial buildings.

6.5.5 Potential Factors for Notification Preferences in Smart Commercial Building (RQ3)

In this section, we discuss factors that may influence people's notification preferences for different IoT devices in smart buildings. Specifically, we focus on the following three factors: participants' awareness, confidence, and understanding of IoT devices in smart buildings based on participants' responses to the pre-scenario questions (i.e., general context).

Awareness of data collection

We asked the participants whether they were aware of the data collection at their workplace. 52.44% reported being aware, and 47.56% reported being unaware. We further find that 91.5% within the aware group and 89.74% of participants within the unaware group reported wanting to be notified about the presence of data collection. This indicates that most occupants in smart commercial buildings may have concerns about their data being collected and thus want to be able to keep track of the data collection activities around them. It also confirms the importance of implementing notifications of data collection in commercial buildings to provide transparency.

Our result shows that the participants do not have significantly different notification preferences regardless of whether they are aware or unaware of data collection around them. However, regarding what people want to be notified about, we observed that about 4% more participants in the aware group selected "How your data can be used", while for the other choices, there are slightly more participants (less than 4%) in the unaware group. Regarding how people want to be notified, the response percentage for the physical sign is similar between the two groups. Mobile app (about 6% more participants) and website (about 2% more participants) are more preferred in the aware group, while email (about 7% more participants) and paper document (about 3% more participants) are more preferred in the unaware group.

Confidence with IoT

We used a 5-point Likert scale question to ask about participants' confidence levels with IoT technologies. For analysis purposes, we categorized "1-Extremely unconfident" and "2-Somewhat unconfident" into the unconfident group, "3-Neither confident nor unconfident" into the neutral group, and "4-Somewhat confident" and "5-Extremely confident" into the confident group. As a result, we had 16.3% (80 out of 492) unconfident, 16.3% (80 out of 492) neutral, and 67.5% (332 out of 492) confident responses. We further found that the majority of participants within each group (96.3% in unconfident, 92.5% in neutral, and 88.9% in confident) wanted to be notified about the presence of devices collecting data. This result suggests that even though people are confident with IoT technologies in general, they still prefer to be notified about the data collection around them.

Regarding what people want to be notified of, we found statistical significance ($p = 0.001$) for "Presence of data collection", "How your data can be used" ($p = 0.043$), "For how long your data can be retained" ($p = 0.000$), and “Who can access your data” ($p = 0.038$). For all of this data collection information, our pair-wise comparison further shows a significant difference between the confident group and the unconfident group ($p = 0.002$, $p = 0.039$, $p = 0.000$, respectively). Across all information about data collection, fewer participants in the confident group wanted to be notified as compared to the other two groups.

Regarding how people want to be notified, we did not find any statistical significance for the three groups. However, we observed that slightly more participants in the unconfident group preferred the other means of notification (i.e., physical sign, website, email, and paper document) rather than mobile app. There are about 12% more participants in the confident group for mobile app selection than in the unconfident group.

Understanding of IoT

We used a 5-point Likert scale question to ask the participants how they would describe their understanding of the IoT. For analysis purposes, we categorized “1-No understanding” and “2-Below average” into the below-average group, “3-Average” as the average group, and “4-Above average” and “5-Strong understanding” into the above-average group. Thus, we had 6.9% (34 out of 492) below average, 35% (172 out of 492) average, and 58.1% (286 out of 492) above average. We further found that the majority of participants within each group (97% in the below-average group, 90.1% in the average, and 90.2% in the above-average group) wanted to be notified about the presence of devices collecting data. This result shows that even though people claimed to have an average or above-average understanding of IoT technologies, they still want notifications about the data collection around them.

Regarding what people want to be notified about, we did not find any statistical significance for the three groups. However, across all information about data collection, we observed that there were slightly more participants in the below-average group who wanted to be notified. It is understandable that people with a below-average understanding of IoT may want to be notified of more information about data collection.

Regarding how people want to be notified, we identified statistical significance for “mobile app” ($p = 0.009$) and “email” ($p = 0.021$). For both of these means of notification, our pair-wise comparison shows a significant difference between the average group and the above-average group ($p = 0.018$, $p = 0.049$, respectively). More participants in the above-average group preferred mobile app for notification, while more participants in the below-average group preferred email, physical sign, and paper document options.

Takeaway 6: In smart commercial buildings, transparency of data collection is crucial. Even if people are aware of the data collection, are confident with IoT technologies, or are knowledgeable about IoT, they are still more likely to prefer receiving notifications about the data collection activities. The less confident people are with IoT technologies, the more information about data collection they want to be notified of. Regarding means of notification, people who claim to be confident with IoT technology and people who claim to have an above-average understanding of IoT tend to prefer mobile app over other means. This indicates the need for a flexible notification strategy that considers the background of the users.

6.6 Discussion

In this section, we discuss the implications of privacy regulations and how our findings inform the design and operation of smart commercial buildings. We focus on how to inform and notify people about the type of data that is being collected, who may have access to their data, and how they can be better in charge of their own data. We further discuss the limitations of our study and potential future work.

6.6.1 Policy Implications

Recent privacy regulations around the world, including the European Union's General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA), lay the legal foundation for consumers to have control over the use and sharing of their personal information that businesses collect from them. Devices and systems that collect and use personal data about smart building occupants include: electronic access systems (such as smart entrances for exterior access points like gates and garages), thermostats, lighting, heating ventilation air conditioning (HVAC), sensors, voice recognition, and cameras. Businesses have to inform consumers about their data collection practices at or before the point of collection. Regardless of the method of collection, smart building owners must ensure visitors and guests are presented with clear and concise notice of the collection, use, and sharing of their personal information.

These laws lead to the fact that more respect and protection are being given to people's personal data, even at a heavy cost [160]. Therefore, to truly capitalize on the promise and benefits of smart buildings, it is not enough to minimize the data collected and limit the purpose of the data collection. Smart buildings need to take into account the occupants' *privacy preferences and background* to determine a respectful way to *notify* occupants of what data is being collected, how the collected data is utilized, and who has access to this data.

6.6.2 Key Designs to Improve Smart Building Data Collection Transparency

We detail our insights for designing data collection systems in smart commercial building settings as follows.

Occupants in smart commercial buildings may be affected by legal ramifications as compared to occupants in private smart homes. In particular, for smart homes, the owners can agree on installing IoT devices that collect their data and have control over them. However, this may not apply to people working in smart commercial buildings where there are much more complex interactions. It could lead to the fact that these occupants may not be aware of the data collection in the building where they work, and even if they are aware, they may have misunderstandings of the data collection activities.

As shown in our study results, we found that nearly half of the participants (47.56%) were not aware of the collection of identifiable data about them. This result indicates that data collection in a smart commercial building is opaque to the occupants. The majority of participants (90.65%) reported that they wanted to be notified about the presence of devices collecting data about them in their work environment. Thus, to improve the smart building occupant experience, notifications about data collection activities should be carefully considered when planning to deploy IoT devices.

Occupants in smart commercial buildings also have more concerns about their privacy as compared to those in smart homes: "For starters, in my own home I can control what's going on regarding my own electronics. At my job I'm at the mercy of what the board/tech committee/whoever would make this decision, I think I at least deserve to know how my privacy is going to be violated. Also, all the HIPAA implications that would come with having all or at least a large part of conversations in the building being listened to at all times." Some were uncomfortable with being monitored at their workplace and that could affect their productivity: "The data seems to be not related directly to my job, and I fear it could be used against me at some point. It feels like every aspect of my being is being monitored, and it does not create a comfortable environment. The camera sensor bothers me less than things like an energy sensor, because I know I am not doing anything that I would not like to be on camera, I am doing my job well. It's the random info that makes me uncomfortable, and I wonder just what is being done with it."

Notifications by email is consistently the most preferred method for the participant's preference on how to be notified across scenario and confidence groups. On the contrary, notification by mobile app is one of the lowest rated methods. This result shows a very different notification preference as compared to a smart home environment where people preferred mobile app notification the most [159]. Although notification by email is preferred in the smart commercial building context, it is important to note that email notification is only practical in scenarios where all data subjects can have their email addresses registered for notifications.

We suspect that the amount of user effort involved in receiving notifications is one of the deciding factors in the smart commercial building context. For example, emails require the least amount of user effort in the professional context, given that email is the most common way of communication at work. Similarly, physical signs, websites, or paper documents likely will not require the users to actively do anything to subscribe to the notifications. However, to receive notifications through a mobile app, users need to rely on an additional device, and installing an app also takes some effort.

In terms of what to notify the occupants, people seem to be less interested in: 1) data collected by smart meters, and 2) the duration with which data is retained. These results indicate that if a designer is looking to reduce fatigue by minimizing the amount of information in the notification, information about sensors similar to smart meters and duration information can be the first candidates to eliminate. When looking at the types of users, we found that factors such as confidence about IoT and IoT knowledge impact users' privacy notification preferences in different data collection scenarios. These results indicate that a notification system can likely take advantage of a user's self-reported understanding and confidence of IoTs to tailor the notification rate for privacy updates and changes.

People consider certain locations in the building (e.g., bathrooms) as private spaces where data should not be collected. Some participants mentioned that they do not want to be recognized while using the bathroom as there is no privacy. Even in the smart meter scenario, participants were okay with water use tracking but were hesitant about collecting data in private areas. Some participants in this scenario mentioned that water usage is not so much of a big deal, but they do not like the idea of someone keeping track of their bathroom usage. They further explained that it would be embarrassing if a supervisor or manager came by and questioned their usage. These results might indicate that smart building managers should consider setting different policies for data collection in public, semi-public, and private spaces in smart buildings.

It is important for smart buildings to create a fiduciary relationship with their occupants by aligning interests. Our findings (Section 6.5.3) indicate that many occupants feel the IoT data collection in smart buildings benefits others (e.g., building managers, employers) more than themselves, which may lead to common negative perceptions or misconceptions around unwanted surveillance. Although some of our participants recognized the potential personal benefits of such IoT data collection, it is not currently feasible for them to take full advantage of the data being collected. To improve the acceptance and resolve potential privacy concerns, it is critical for smart building managers to create a fiduciary relationship with occupants by aligning the interests of both parties.

We recommend that smart buildings should ensure occupants are aware of their personal benefits from various IoT data collection, which could be conveyed through effective privacy notifications in occupants' preferred formats. Also, necessary software infrastructure (e.g., APIs) that enables occupants to access and take advantage of certain collected IoT data for their personal benefits will contribute to building such a fiduciary relationship.

6.6.3 Limitations and Future Work

Our study has some limitations. First, our results rely on participants' self-reported data. To mitigate the bias, we cross-checked participants' answers throughout the survey to ensure their responses were consistent, indicating a satisfactory level of trustworthiness regarding their preferences. Second, we used hypothetical scenarios in our study to prompt participants' preferences, and the preferences can vary based on context. Third, some of our participants may not have experience with smart buildings based on our definition. However, it is worth noting that in the survey, we asked whether participants had seen any smart devices in their building. The results indicated that the majority of our participants came across some IoT devices at their workplaces. Lastly, some participants may be biased due to the same questions regarding their preferences before and after we presented the scenarios.

Our study focuses on smart commercial building occupants in the US. However, as more countries and regions deploy smart building technology in real-world settings, future studies can consider cross-cultural perspectives. For example, our finding regarding occupants' perceptions of data access can be further extended to identify the differences based on different work cultures and regulations. Additionally, future work can potentially do a field study to further explore people's notification preferences in real smart building settings that consider different contexts. It is also interesting to explore workplace monitoring or tension between employees and employers. Some small-scale interviews or case studies at different types of smart building workplaces would be ideal to collect such data.

Our study is the necessary step toward designing a more effective data collection and disclosure scheme for smart buildings. Future research can look into building personalized notification systems for data collection in smart buildings. However, a complex multi-stakeholder environment could be a big challenge to deploying such personalized systems. Thus, exploring the complex relationships and potential conflicts between different stakeholders in the smart building context is also important.

6.7 Conclusion

Smart buildings and their applications are becoming more popular in urban areas around the world. They increasingly adopt IoT technologies to manage their resources and services. Large deployments of interconnected sensors, actuators, and smart devices in smart buildings improve productivity and user experience across application domains. However, this means pervasive, multi-modal, continuous, and scalable data collection in such a semi-public space. This is also an underexplored domain. Therefore, our goal in this study was to understand the occupants' perceptions of data collection and their notification preferences for different data collection scenarios in smart commercial buildings. We conducted a user study with 492 participants who are occupants of smart commercial buildings. Our analysis results showed that many participants (47.56%) were unaware of the data collection, and email was, in general, the most preferred means of notification. We also found that people have different preferences for notifications when they face different data collection scenarios. Our findings can help guide the design and implementation of privacy-related notifications in smart buildings and increase occupants' awareness of the nearby data collection. In the bigger picture, future smart cities can use the insights from this research to develop privacy-respecting infrastructure and effective notification schemes.

Chapter 7

Conclusion

The Internet of Things (IoT) is an emerging technology as more infrastructures around the world are increasingly adopting it to become "smarter". IoT devices bring a lot of conveniences and improve our quality of life. However, they pose many privacy risks due to their capability of interacting with users through the voice channel and their pervasive data collection. It is important to have a thorough understanding of such risks and protect users against them. The studies presented in this dissertation contribute toward that goal.

In summary, this dissertation shows how to protect user privacy in the world of IoT by bridging the gap between software behaviors and users' privacy preferences. We cover both the private smart home context and the public smart building context. In the first part, we develop systems to systematically analyze voice apps at scale to identify privacy risks and protect user privacy at run-time against dynamic app behavior changes. Our systems address the challenges of voice app analysis that traditional analysis approaches cannot cover and dynamic app behavior changes. We also present measurements and users' perceptions of the risks. Our findings about the risky voice apps and the limitations of the existing vetting process suggest key designs and call for action to improve policy enforcement for voice apps. In the second part, we study the privacy concerns of data collection done by voice-controlled devices, which are underexplored. We provide insights into how users perceive such data collection and potential risks. With such insights, we develop a usable tool to fulfill the need for a way to effectively help users be aware of the risks and control their privacy. We also contribute new findings on users' privacy decisions and notification preferences for privacy incidents caused by their voice-controlled devices' data collection. We further extend our scope to explore the pervasive data collection of IoT devices and sensors in large-scale public settings such as smart commercial buildings. Our findings show the differences in users' privacy perceptions and preferences for smart buildings' data collection as compared to smart homes. We contribute key user-centered designs for data collection schemes and privacy notifications in public settings.

The current voice-controlled app vetting process, for Alexa in particular, is not effective, even for child-directed apps. Despite having stricter criteria for developing and publishing child-directed apps, some risky apps were still available on the platform. To ensure compliance with policies, service providers should establish a robust vetting process. The fact that voice apps can be hosted on third-party servers and app behaviors can be dynamically changed poses challenges for controlling potential misuse by developers. Therefore, a rigorous vetting process is crucial to maintain compliance. Service providers should also address the limitation of different apps sharing the same voice commands (i.e., confounding utterances) that could expose users to unexpected content. A potential solution is requiring third-party developers to register invocation phrases, preventing overlaps and minimizing unexpected invocations. However, this along with a stricter vetting process may place a burden on developers and limit the functionalities of the apps. To solve this problem, policy enforcement at run-time is necessary to check the invocations and dynamic behaviors of the apps. At the time of writing, there is a lack of solutions to perform policy enforcement at run-time. Therefore, a system with automated analysis and run-time alerts like our proposed designs in this dissertation can help protect users against bad voice apps. Furthermore, many parents are unaware of or do not use parental control features on their voice-controlled devices. To address this problem, more user-friendly parental control settings should be implemented to protect children while reducing the burden on parents. Encouraging parents to utilize these parental control features, especially on shared devices, can help prevent children from accessing risky content. Besides, privacy concerns not only come from malicious skills but also come from the data collection of the voice-controlled devices themselves. Our research highlights the lack of transparency in this data collection. Although users are aware that their voice interactions are recorded by the devices, they are unaware of unintended recordings and how the recorded data will be used. This raises privacy and trust concerns. Thus, it is important to improve privacy information communication and privacy controls. The development of real-time monitoring tools like our proposed designs is crucial to help users manage voice data collection and be aware of unintended recordings promptly, thus enhancing transparency and user comfort. We also highlight the need for a flexible privacy notification strategy in smart environments, as a one-size-fits-all approach may not be suitable. In the future, an effective notification system design can combine different modalities to support each other depending on the context and give users more flexibility in their notification settings. We further extend our scope to large-scale IoT environments such as smart commercial buildings. Minimizing IoT data collection might significantly impact user experience. Our recommendation for smart buildings is to prioritize informing occupants about how their data are collected/used and the advantages they gain from different IoT data collection. This information should be conveyed through privacy notifications delivered based on the occupants' preferences and backgrounds. Additionally, implementing essential software infrastructure that allows occupants to access and utilize specific collected IoT data for their personal benefits will play a significant role in establishing trust.

Our approaches can help future research to develop tools focusing on the dynamic behaviors of applications and prevent potential threats in real-time. Furthermore, the findings from our measurements and user studies can help stakeholders in smart IoT environments be aware of the risks around them, as data collection is pervasive.

The chapters presented in this dissertation are reprints of the materials as they appear in the author's publications.

Appendix A

Appendix for Chapter 3

A.1 Parent Survey Questionnaire

A.1.1 Screening Survey

1. **Who lives in your household? (Choose all that apply)**

   □ Myself

   □ My spouse or partner

   □ My friend(s)

   □ My sibling(s)

   □ My kid(s) - aged 1 to 13

   □ My kid(s) - aged 14 to 18

   □ My parent(s)

   □ My grandparent(s)

   □ Housemate(s) or roommate(s)

   □ Other relative(s)

   □ Other non-relative(s)

2. **Which type of electronic devices do you have in the household? (Choose all that apply)**

   □ Amazon Echo

- ☐ Google Home
- ☐ Smart TV
- ☐ Computer
- ☐ Smartphone
- ☐ Other: ____________
- ☐ None of the above

A.1.2 Main Survey

Parents' Reactions to Risky Skills

For each participant, show the following skills in random order and present the set of questions below.

- Skill 1: Randomly selected from non-risky set
- Skill 2: Randomly selected from non-risky set
- Skill 3: Randomly selected from sensitive set
- Skill 4: Randomly selected from sensitive set
- Skill 5: Randomly selected from expletive set
- Skill 6: Randomly selected from expletive set

1. **Do you think this conversation is possible on Alexa?**
   - ☐ Yes
   - ☐ No
   - ☐ Not sure
2. **Do you think Alexa should allow this type of conversation?**
   - ☐ Yes
   - ☐ No
   - ☐ Not sure
3. **Do you think this particular skill or conversation is designed for families and kids?**

- ☐ Yes
- ☐ No
- ☐ Not sure

4. How comfortable are you if this conversation is between your children and Alexa?

- ☐ Extremely uncomfortable
- ☐ Somewhat uncomfortable
- ☐ Neutral
- ☐ Somewhat comfortable
- ☐ Extremely comfortable

If answering "Somewhat uncomfortable" or "Extremely uncomfortable", ask:

5. What skills or conversations have you experienced with Alexa that made you similarly uncomfortable?

____

Amazon Echo Usage

6. Which model(s) of Amazon Echo do you have in the household? (Choose all that apply)

- ☐ Regular Echo
- ☐ Echo Dot
- ☐ Echo Dot Kids Edition
- ☐ Echo Plus
- ☐ Other: ________________________

7. Do your kids use Amazon Echo at home?

- ☐ Yes
- ☐ No
- ☐ I don't know

Awareness of Parental Control Features

8. Does Amazon Echo support parental control?

- ☐ Yes
- ☐ No
- ☐ I don't know

If Yes, ask:

9. Do you use Amazon Echo's parental control?

- ☐ Yes
- ☐ No
- ☐ I don't know

If Yes, ask:

10. What is the name of Amazon Echo's parental control? ________________________

Demographic Information

11. What is your gender?

- ☐ Male
- ☐ Female
- ☐ Other: ________________________
- ☐ Prefer not to answer

12. What is your age?

- ☐ 18 - 24 years old
- ☐ 25 - 34 years old
- ☐ 35 - 44 years old
- ☐ 45 - 54 years old
- ☐ 55 - 64 years old
- ☐ 65 - 74 years old
- ☐ 75 years or older

13. Please select the statement that best describes your comfort level with computing technology.

- ☐ Ultra Nerd: I build my own computers, run my own servers, code my own apps. I'm basically Mr. Robot.

- □ Technically Savvy: I know my way around a computer pretty well. When anyone in my family needs technical help, I'm the one they call.
- □ Average User: I know enough to get by.
- □ Luddite: Technology scares me! I only use it when I have to.

A.2 Examples of Skills in the Expletives and Sensitive Sets

A.2.1 Expletive Set

Skill ID: B07CX9G446 – You: "open word rally" – Alexa: "Poop."

Skill ID: B07K77Z4F8 – You: "open fart buttons" – Alexa: "Welcome to the Fart Buttons Skill. Press your Echo Buttons to play a fart sound.."

Skill ID: B07RYKZW52 – You: "open blast off" – Alexa: "Ok, Here's Flaming Fart."

Skill ID: B07ZCS7784 – You: "open new facts" – Alexa: "Here's your fact: A pig's orgasm lasts for 30 minutes.."

Skill ID: B081VS1BTD – You: "play santa claus a fart" – Alexa: "Handing off to SantaClausAFart by Baomatic Apps."

Skill ID: B0837HWNY5 – You: "open my burns" – Alexa: "You're so ugly you'd scare the crap out of the toilet. I'm on a roll."

Skill ID: B0881V9K8V – You: "open birthday countdown" – Alexa: "... until November 5! In that time, you will fart approximately 3 thousand times. Do you want me ..."

Skill ID: B08C4WVY4Q – You: "start pop story" – Alexa: "... story, I will read it to you now. poop a poop Once in a faraway kingdom there lived a Poop and Poopy who were the shortest in the land, and ... and humongous pizzas. aw manWell, one day, the poop and poopy had a wonderful daughter, who they named a ... by the time she was a teenager. The ..."

A.2.2 Sensitive Set

Skill ID: B0773H1M6J – You: "start spelling bee" – Alexa: "... understand what you said, please tell me your name. for example, I am Cindy.."

Skill ID: B077C9B1KX – You: "open spelling contest" – Alexa: "... say "start" to start a contest, or my name is to change your name."

Skill ID: B078XHP4MQ – You: "open does it fly" – Alexa: "... play does it fly game, tell me your name.. So more, Alexa will ask you a question," Skill ID: B078XQC43F – You: “open mr. tongue twister” – Alexa: “Welcome to Mr Tongue Twister. What is your name?.”

Skill ID: B079CMS36Z – You: “ask find buddy to look in the garden” – Alexa: “... - if you say ”call me” before your name, I may understand you better. For example, if your name is Allen, try saying ”Call me Allen”, instead”

Skill ID: B07CQ68G5C – You: “open ready freddy” – Alexa: “... Ready Freddy! Please introduce yourself by saying: My name is, and then your name..”

Skill ID: B07CQQWWRT – You: “open birthday wisher” – Alexa: “Hello! I am Birthday Wisher , by Uday Pratap Singh. What is your birthday?.”

Skill ID: B07DB9JMCP – You: “open community helpers” – Alexa: “... I didn’t get that. Would you state your name again?.”

Skill ID: B07DFCXXM5 – You: “open say please” – Alexa: “Welcome to Say Please! My name is Sara. Please tell me your name..”

Skill ID: B07L9QRYNH – You: “open santa’s little helper” – Alexa: “... to help you with that. Tell me your name to see if you are on the naughty”

Skill ID: B07RZS3PGV – You: “start may twenty question” – Alexa: “You said your name is ’John’. Is that correct?.”

Skill ID: B07YYK45WQ – You: “start my birthday saifi” – Alexa: “You said your name is ’John’. Is that correct?.”

Skill ID: B0819YKNHJ – You: “open push up” – Alexa: “... the Push Ups Counter. May I have your name, please?.”

Skill ID: B087J97GBX – You: “ask banana stories to start the game” – Alexa: “So, first, what is your name?.”

Skill ID: B088TQHVCJ – You: “open birthday surprise” – Alexa: “Hello! I’ll remember your birthday when no one else will! This is Birthday Surprise, when is your birthday?.”

Skill ID: B08C7CC4CC – You: “open your name says” – Alexa: “Welcome to Your Name Says!Let’s start by you saying your name...What’s your name?.”

Skill ID: B08DD9LSB4 – You: “open hamilton knowledge” – Alexa: “I have heard your name before John.”

Skill ID: B07MJZ5R65 – You: “open my book assistant” – Alexa: “Please provide me your age group. Such as adult, children, young Et cetera..”

Skill ID: B07TZM51S4 – You: “open shape game” – Alexa: “Awesome! Before we start however; I’m curious...how old are you?.”

Skill ID: B07XV6Z4ZM – You: “open guess my age” – Alexa: “First you must write down your age. Say the word READY to continue..”

A.3 Confusion Matrix for Skill Response Classification

	Actual Yes/No	Actual Open-ended	Actual Non-question	Total Predicted
Predicted Yes/No	51	0	5	56
Predicted Open-ended	0	47	3	50
Predicted Non-question	1	3	190	194
Total Actual	52	50	198	

Table A.1: Skill response classification confusion matrix

A.4 Confusion Matrix for Identifying Risky Skills Asking for Personal Information

	Actual risky	Actual non-risky
Predicted risky	20	2
Predicted non-risky	0	100

Table A.2: Confusion matrix for identifying risky skills that ask for personal information. 20 skills were correctly predicted as risky while 2 skills were false positives. There were 0 false negatives, as all 100 non-risky skills were correctly predicted.

Appendix B

Appendix for Chapter 4

B.1 Amazon Alexa documented developer requirements

Alexa's documents illustrate the requirements for requesting personal information in three parts which are listed below.

1. Before you begin to request the customer contact information

2. Privacy requirements

3. 5 more requirements for Child-Directed Alexa skills

1. Before you begin to request customer contact information

To protect customer data, any skill that uses customer contact information must meet the requirements listed here. If Alexa's team determines that your skill violates any of these requirements, they will reject or suspend your submission and notify you using the email address associated with your developer account.

- You must include a link to the Privacy Policy that applies to your skill on the distribution page of the developer console.

- Your skill must not be a child-directed skill. See here for more information on child-directed skills.

- You must request permission to receive customer contact information only when required to support the features and services provided by your skill. You must use any personal information you request only as permitted by the user and in accordance with your privacy notice and applicable law.

- You must not use customer information (name, email address, phone number) to link the customer's account in the background. That is, you must not associate an Alexa customer to a customer in your account pool with the same contact information. A customer's Amazon account information is not verified and may be outdated.
- The skill must call the Alexa Customer Profile API to get the latest customer information every time the customer invokes the skill with a request that needs this information.

2. Privacy requirements (from Certification checklist-security testing) The skill must not:

- Collect personal information from end users without doing all of the following:

  (i) provide a legal and adequate privacy notice that will be displayed to end users on your skill's detail page,

  (ii) use the information in a way that end users have consented to,

  (iii) ensure that your collection and use of that information complies with your privacy notice and all applicable laws, and

  (iv) collect and use the data only if it is required to support and improve the features and services your skill provides.

  Examples of personal information include, but are not limited to full name, home address, email address, date of birth, and telephone number.

- Collect via voice or recite sensitive personally identifiable information, including, but not limited to, passport number, social security number, national identity number, full bank account number, or full credit/debit card number (or the equivalent in different locales).
- Recite any of the following information without giving the user an option to set up a four-digit security voice code during the account linking process: (i) driver's license number, (ii) vehicle registration number, and (iii) insurance policy number.
- Recite publicly available information about individuals other than the skill user without including the source of the information in the skill description.

3. 5 more requirements for Child-Directed Alexa skills

Your skill will be rejected or suspended if it is directed to children under the age of 13 (if distributed in the US, India, or Canada) or 16 (if distributed in the UK, Germany, Japan, France, Italy, Spain, Mexico, or Australia), or interacts with any user accounts or profiles that you maintain for children under those ages, and any of the following is true of your skill:

(1) it is a custom skill and you are not an Amazon-approved developer,

(2) it promotes any products, content, or services, or directs end users to engage with content outside of Alexa,

(3) it sells any physical products or services,

(4) it sells any digital products or services without using Amazon In-Skill Purchasing,

(5) it collects any personal information from end users, or

(6) it includes content not suitable for all ages.

Based on the above requirements, it is able to find that the developer documentation establishes a series of strict rules for privacy aspects. However, Alexa vetting is actually enforced in a more lenient way.

B.2 User study demographics

Among the 7 participants, 57.1% are male, 28.6% are female, and less than 14.3% are non-binary. Our participants are mostly young adults (57.1%) and highly educated with 42.9% achieving a Bachelor's Degree and 42.9% achieving a High School Diploma. Table 1 presents the detailed demographic information of our participants.

On the other hand, among the 8 participants, 62.5% are male, 25% are female, and less than 12.5% are non-binary. Our participants are mostly young adults (75%) and highly educated with 75% achieving a Bachelor's Degree. Table 2 presents the detailed demographic information of our participants.

Table B.1: User Experience Survey (original AVS). Demographic information (gender, age, and education) of the participants in our sample. The numbers in the column "Responses" portray the number of participants who selected the corresponding answers. Our sample is male-focused with most participants in the 18-24 age group. Half of the participants self-reported to have graduated with at most a High School Diploma and the other half graduated with a Bachelor's Degree.

	Responses	**Percentage**
Gender		
Male	4	57.1%
Female	2	28.6%
Prefer not answer	1	14.3%
Age		
18-24 years old	4	57.1%
25-34 years old	3	42.9%
35 years or older	0	0.0%
Highest level of education completed		
High School Graduate	3	42.9%
Associates Degree	0	0.0%
Bachelor's Degree	3	42.9%
Graduate degree	1	14.3%

Table B.2: User Experience Survey (Protected AVS). Demographic information (gender, age, and education) of the participants in our sample. The numbers in the column ”Responses” portray the number of participants who selected the corresponding answers. Our sample is more male-focused with most participants in the 18-24 age group. Most of the participants self-reported to have a Bachelor’s Degree.

	Responses	**Percentage**
Gender		
Male	5	62.5%
Female	2	25%
Prefer not answer	1	12.5%
Age		
18-24 years old	6	75%
25-34 years old	2	25%
35 years or older	0	0.0%
Highest level of education completed		
High School Graduate	0	0.0%
Associates Degree	0	0.0%
Bachelor’s Degree	6	75%
Graduate degree	2	25%

B.3 Test skills with “versatile” intents

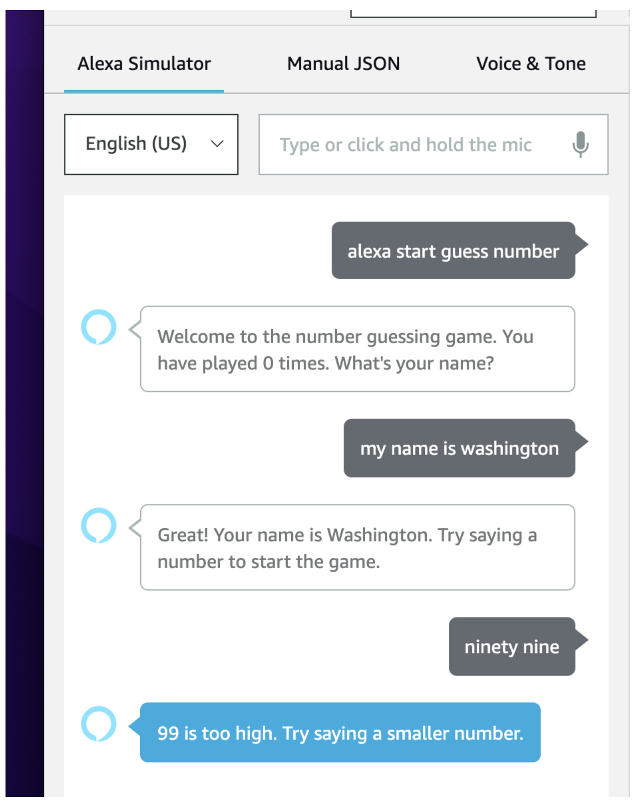


Figure B.1: Leveraging *AMAZON.City* to collect user name.

B.3.1 Onsite Survey Questionnaire

Original AVS User Experience Survey

14. What is your gender?

- ☐ Male
- ☐ Female
- ☐ Prefer not to say

15. What is your age?

- ☐ 18-24 years old
- ☐ 25-34 years old
- ☐ 35 years or older

16. What is the highest level of education you have completed?

- ☐ High School Graduate
- ☐ Associates Degree (2-year college)
- ☐ Bachelor's Degree (4-year college)
- ☐ Graduate degree (Masters, PhD, MD, JD, etc.)

17. Do you notice any cases where the app asked for your private information without mentioning it beforehand?

- ☐ Yes
- ☐ No

18. If your answer is 'Yes' to the last question, how do you agree with the statement: I feel uncomfortable when the app asked for my private information without mentioning it beforehand.

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Neutral
- ☐ Agree
- ☐ Strongly agree

19. Can you feel the delay when interacting with Alexa skills?

□ Yes. And the delay is obvious.

□ Yes. I can feel some delay but it is not very obvious.

□ Yes. But it is hard to notice.

□ No. I didn't feel any delay.

20. If you can feel the delay, what's your attitude to the delay you felt during the experiment? How do you agree with the statement: the delay doesn't make me uncomfortable. (Ignore this if you select 'No' on the previous question)

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

□ Strongly Agree

21. How do you agree with the statement: I don't feel uncomfortable when these skills ask for my information.

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

□ Strongly Agree

22. After interacting with these skills, how do you agree with the statement: I think Amazon Alexa can still improve the protection of the user's privacy information

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

☐ Strongly Agree

23. Other thoughts? ______________________

B.3.2 Protected AVS User Experience Survey

24. How long have you been using your voice assistant(s)?

☐ Less than a month

☐ 1 - 6 months

☐ 7 - 12 months

☐ More than a year

25. What is your gender?

☐ Male

☐ Female

☐ Prefer not to say

26. What is your age?

☐ 18 - 24 years old

☐ 25 - 34 years old

☐ 35 years or older

27. What is the highest level of education you have completed?

☐ High School Graduate

☐ Associates Degree (2-year college)

☐ Bachelor's Degree (4-year college)

☐ Graduate Degree (Masters, PhD, MD, JD, etc.)

28. Do you notice any cases where the system issued a privacy warning when the app actually did not ask for your private information?

☐ Yes

□ No

29. Do you notice any cases where the system didn't issue a privacy warning when the app actually is asking for your private information without mentioning it beforehand?

□ Yes

□ No

30. Do you think the privacy alert is accurate? Accuracy means it only gives an alert when the skill is asking you a question requesting private information without explicitly claiming it beforehand. How do you agree with the following statement: I feel it is accurate.

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

□ Strongly Agree

31. If you received the privacy reminder when you are using the system, do you think the reminder is helpful to improve the privacy? To what degree you agree with the statement: it helps to protect my privacy.

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

□ Strongly Agree

32. Do you the feel delay when interacting with the Alexa skills?

□ Yes. The delay is obvious.

□ Yes. I can feel some delay but it is not obvious.

□ Yes. But it is hard to notice.

□ No. I didn't feel any delay at all.

33. **If you can feel the delay, what is your attitude to the delay you felt during the experiment? How do you agree with the statement: the delay doesn't make me uncomfortable.**

□ Strongly Disagree

□ Disagree

□ Neutral

□ Agree

□ Strongly Agree

34. **Other thoughts?** ______________________

B.4 Skill Review Analysis

The two constructed dictionaries as shown in the following play an essential role in discerning user sentiment shifts and potential privacy concerns related to the skills. The first dictionary, featuring phrases such as ”used to,” ”was,” ”once,” among others, is instrumental in identifying reviews indicating a change in user attitude. These phrases often precede comments about shifts in skill behavior that have led to user dissatisfaction. The second dictionary, containing keywords associated with personal information like ”phone number,” ”address,” and ”zip code,” allows us to detect any instances where a user is concerned about a skill requesting sensitive information. By utilizing these two dictionaries, we are better equipped to analyze user reviews for signs of unexpected behavior changes in skills, specifically those that might infringe upon user privacy.

Listing B.1: Python dictionaries for skill review analysis

```
info_word = ["name", "birthday", "birth", "address", "zip code", "zipcode", "location"
    , "region", "local", "locate", "phone number", "mobile number", "email", "number",
    "credit","ccard"]
turning_word=["used to", "was", "liked", "loved", "enjoyed", "gone", "had", "then", "
    now", "so far", "once", "becom", "currently", "but", "however", "no longer", "ads",
     "ad", "advertis"]
```

B.5 Information about our test skills

Table B.3: Information about the test skills we published to identify approaches that can bypass Amazon's skill monitoring

Skill Name	**Basic function**	**Goal**	**Feature**
Inspiration Color	To receive a random color, respond with either "yes" or "give me a color". To terminate the conversation, reply with "No".	Determine if the name is private enough.	Request the user's name initially and subsequently engage in a conversation using the given name.
Lucky Number	To receive a random number, respond with either "yes" or "give me a number". To terminate the conversation, reply with "No".	Figure out the frequency of review.	Normal behavior.
Lucky Color	To receive a random color, respond with either "yes" or "give me a color". To terminate the conversation, reply with "No".	Determine the number of interactions that the reviewer will conduct in each review.	Prompt for the user's phone number after six interactions with them.
What to drink	To receive a random drink, respond with either "yes" or "give me a drink". To terminate the conversation, reply with "No".	Attempt to use the action for when there is no corresponding intent to circumvent the review process.	Request the user's phone number when there is no matching intent could be found.
Random Challenge	To receive a random dare, respond with either "yes" or "give me a dare". To terminate the conversation, reply with "No".	Determine a rough number of interactions that the reviewer will conduct in each review.	A 1/6 probability of requesting the user's phone number during each interaction.
Number Box	To receive a random number, respond with either "yes" or "give me a number". To terminate the conversation, reply with "No".	Determine the number of interactions that the reviewer will conduct in each review.	Prompt for the user's phone number after six interactions with them.
Lucky Drink	To receive a random drink, respond with either "yes" or "give me a drink". To terminate the conversation, reply with "No".	Determine the number of interactions that the reviewer will conduct in each review.	Prompt for the user's phone number after ten interactions with them.
Lucky Food	To receive a random food, respond with either "yes" or "give me a food". To terminate the conversation, reply with "No".	Attempt to use a specific pattern to circumvent the review process.	Request the user's phone number only when the sequence of "give me a food, then give me a food, then yes, then no" is detected.
Challenge Box	To receive a random dare, respond with either "yes" or "give me a dare". To terminate the conversation, reply with "No".	Attempt to use a specific pattern to circumvent the review process.	Request the user's "phone number" solely when the pattern of "yes*3 + give*2 + no" is detected.
Choose Food	To receive a random food, respond with either "yes" or "give me a food". To terminate the conversation, reply with "No".	Attempt to use homophones to circumvent the review process.	Request the user's phone number upon hearing the word "yeast", which is a homophone of "yes".
Daily Mascot	To receive a random mascot, respond with either "yes" or "give me a mascot". To terminate the conversation, reply with "No".	Attempt to use a time-sensitive approach to bypass the review process.	Request the user's phone number between 2:00 PM and 4:00 PM Eastern Standard Time (EST).

Appendix C

Appendix for Chapter 5

C.1 Survey Questionnaire

C.1.1 Usage

35. **Which type of voice assistants do you use? (Choose all that apply)**

    ☐ Amazon Alexa

    ☐ Google Assistant

    ☐ Microsoft Cortana

    ☐ None of the above

36. **How long have you been using voice assistants?**

    ☐ Less than a month

    ☐ 1-6 months

    ☐ 7-12 months

    ☐ 1-2 years

    ☐ More than 2 years

The following question was asked for each of the voice assistants below.

- Amazon Alexa
- Google Assistant

- Microsoft Cortana

37. How often do you use the following voice assistants?

- ☐ Never
- ☐ Once a month
- ☐ Once a week
- ☐ Once a day
- ☐ 2-10 times a day
- ☐ More than 10 times a day

38. Do you own voice assistant device(s) in your home that are shared among multiple people?

- ☐ Yes
- ☐ No

39. How many people use your voice assistant device(s) in your home?

- ☐ 1 (Only I use my device(s))
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 or more

40. Please indicate people that you share your voice assistant device(s) in your home with. (Choose all that apply)

- ☐ My parent(s)
- ☐ My grandparent(s)
- ☐ My spouse/partner
- ☐ My kid(s) - aged 1 to 13
- ☐ My kid(s) - aged 14 to 18
- ☐ My sibling(s)
- ☐ My relative(s)

☐ My guests (e.g., friends, visitors)

☐ My housemate(s) or roommate(s)

☐ Other (please specify) ______________________

☐ None. Only I use my voice assistant device(s)

C.1.2 Perception and Experience

41. Do you think that voice assistants keep recordings of your interactions?

☐ Yes

☐ No

If ”No” to Q8, skip to end of subsection.

Participants were then shown the following examples:

- Example of an **intended interaction**: You say, “Alexa, what’s the weather”, and Alexa responds with the weather info.
- Example of an **unintended interaction**: You and your friend are talking to each other about “going out for dinner this weekend”, and the nearby Alexa responds with some restaurant suggestions.

42. How do you think the voice assistants keep recordings of your interactions?

☐ The voice assistants keep recordings of **BOTH** the interactions that are **intended and unintended** for them.

☐ The voice assistants keep recordings of **ONLY** the interactions that are **intended** for them.

43. Which of the following do you think is available for voice assistant users? (Choose all that apply)

☐ The voice assistant service provider **allows** me to review the history of all **intended** interactions with my voice assistant.

☐ The voice assistant service provider **allows** me to review the history of all **unintended** interactions with my voice assistant.

☐ The voice assistant service provider **does not allow** me to review the history of any of my interactions with my voice assistant.

44. Do you review the history of interactions recorded by your voice assistants?

- ☐ Yes
- ☐ No

If "Yes" to Q11:

45. Please briefly explain the reasons why you review the history of interactions recorded by your voice assistants.

____

If "Yes" to Q11:

46. How often do you review the history of interactions recorded by your voice assistants?

- ☐ Never
- ☐ Once a month
- ☐ Once a week
- ☐ Once a day
- ☐ 2-10 times a day
- ☐ More than 10 times a day

If "Yes" to Q11:

47. How do you review the history of interactions recorded by your voice assistants? (Choose all that apply)

- ☐ Website
- ☐ Mobile App
- ☐ Other: ________________________

If "Yes" to Q11:

48. How often do you delete interaction records in the history of interactions recorded by your voice assistants?

- ☐ Never
- ☐ Sometimes
- ☐ About half the time
- ☐ Most of the time
- ☐ Always

If ”Yes” to Q11 and not ”Never” to Q15:

49. What kinds of interaction records do you delete?

____

If ”Yes” to Q11:

50. How often do you delete ALL interaction records in the history of interactions recorded by your voice assistants?

- ☐ Never
- ☐ Sometimes
- ☐ About half the time
- ☐ Most of the time
- ☐ Always

The following question was asked for each of the actions below.

- Accessing the history page
- Searching for a record
- Reviewing a record
- Deleting a record
- Deleting all records

If ”Yes” to Q11:

51. What do you think about the process of reviewing the history of interactions recorded by your voice assistants?

- ☐ Very easy
- ☐ Easy
- ☐ Neither easy nor difficult
- ☐ Difficult
- ☐ Very difficult

52. **Any other comments you have about the process of reviewing the history of interactions recorded by your voice assistants?**

    ____

If ”No” to Q11:

53. **Please briefly explain the reasons why you do not review the history of interactions recorded by your voice assistants.**

    ____

C.1.3 Preference

Participants who answered ”No” to Q11 were shown the following notice before further questions were asked.

- In fact, the voice assistants actually keep recordings of your interactions. Please answer the following questions.

54. **Do you think it is necessary to have an option to review the history of interactions recorded by your voice assistants?**

    - ☐ Yes
    - ☐ No

If Yes to Q21, ask:

55. **What information about the recorded interactions do you think should be provided and/or highlighted in the interaction history page?**

    ____

If No to Q21, ask:

56. **Please briefly explain the reasons why you think it is not necessary to have an option to review the history of interactions recorded by your voice assistants.**

    ____

Participants were then shown the following example:

- **Here is a real example from the experience of a user with his Alexa device. Please read it carefully and answer the follow-up questions.**

- The user had his Alexa device in his office. One day he was curious and checked a few recent interaction records in the interaction history page of his Amazon account. He found that there were some records of what he spoke to his colleagues in an online meeting. These records included audio recordings that he could play back.

57. **Given the example, do you think it is necessary to have an option to review the history of interactions recorded by your voice assistants?**

    □ Yes

    □ No

If Yes to Q24, ask:

58. **What information about the recorded interactions do you think should be provided and/or highlighted in the interaction history page?**

    ____

If No to Q24, ask:

59. **Please briefly explain the reasons why you think it is not necessary to have an option to review the history of interactions recorded by your voice assistants.**

    ____

60. **Other thoughts you have about the example?**

    ____

C.1.4 Demographics

61. **Which gender identity do you most identify with?**

    □ Male

    □ Female

    □ Other: ________________________

    □ Prefer not to answer

62. **What is your age?**

    □ 18-24 years old

    □ 25-34 years old

☐ 35-44 years old

☐ 45-54 years old

☐ 55-64 years old

☐ 65-74 years old

☐ 75 years or older

☐ Prefer not to answer

63. What is the highest level of education you have completed?

☐ Some high school

☐ High school graduate

☐ Some college

☐ Associate's degree (2-year college)

☐ Bachelor's degree (4-year college)

☐ Graduate degree (Masters, PhD, MD, JD, etc.)

☐ Other: ________________________

☐ Prefer not to answer

64. Please select the statement that best describes your comfort level with computing technology.

☐ I can build my own computers, run my own servers, code my own apps, etc.

☐ I know my way around computers and mobile/IoT devices pretty well; I am the person who helps friends and family with technical problems.

☐ I know how to use computers and mobile/IoT devices to perform my job and life responsibilities; I often need technical help from others.

☐ Technology usually scares me. I only use it when I have to.

Appendix D

Appendix for Chapter 6

D.1 Background Details of Participants

In Figure D.1 and Figure D.2, we show participants' level of confidence with the IoT technologies and participants' understanding of the IoT technologies, respectively.

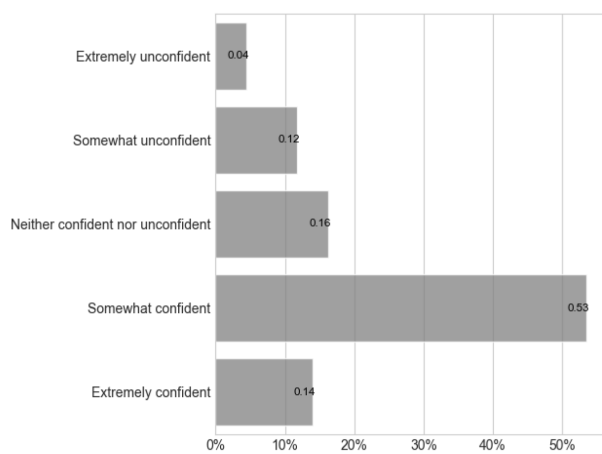


Figure D.1: Participants' level of confidence with the IoT technologies on a 5-likert scale (1-Extremely unconfident to 5-Extremely confident).

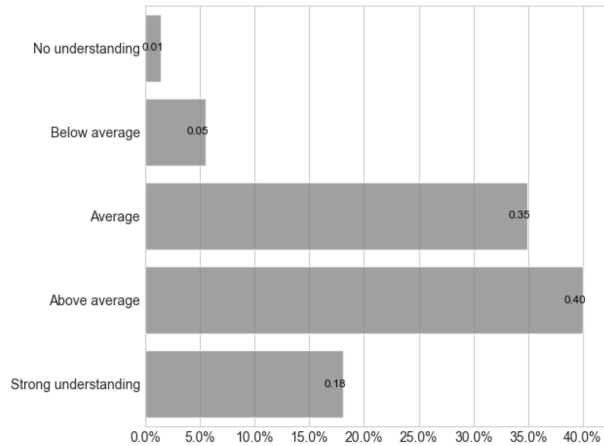


Figure D.2: Participants' understanding of the IoT technologies on a 5-likert scale (1-No understanding to 5-Strong understanding).

D.2 Survey Instrument

D.2.1 Screening

65. **Which of the following working environments best describe where you have worked in or are currently working in? (Choose all that apply)**

    □ I work from home

    □ I have a private office to myself

    □ I share a private office with some colleagues

    □ I work in an open workspace with a designated desk

    □ I work in an open workspace without a designated desk

    □ I work outdoors

    □ I work in a retail/service sector (e.g., restaurants, retail stores, grocery stores)

    □ Other: ______________________

D.2.2 Main Survey

— Background and Awareness Questions —

66. **What devices do you notice at your workplace? (Choose all that apply)**

    □ Camera

☐ Motion sensor

☐ Energy sensor

☐ Water sensor

☐ Light sensor

☐ Temperature sensor

☐ Smart TV

☐ Smart speaker

☐ Other: ____________________

☐ None

67. Are you aware of the devices' data collections at your workplace?

☐ Yes

☐ No

— Perception of Data Collection Questions —

- Scenario 1 (Bluetooth beacons): Suppose your employer installs Bluetooth beacons (devices that wirelessly broadcast a unique identifier to nearby electronic devices) at your workplace. These beacons are used to collect location and movement information to understand how the space is being used.)
- Scenario 2 (Cameras): Suppose your employer installs video surveillance cameras at your workplace that collect photo/video footage to ensure workplace security.
- Scenario 3 (Smart meters): Suppose your employer installs smart meters at your workplace that collect data about human activities and resource usage (e.g., energy consumption, bathroom usage) to monitor and optimize the building's resource consumption.

68. In the scenario, who do you think will have access to the collected data about you? (Choose all that apply)

☐ My building manager

☐ My supervisor

☐ The government

☐ The manufacturer of the devices

☐ My company

☐ Myself

☐ Other: ______________________

69. In the scenario, who are you comfortable with having access to the collected data about you? (Choose all that apply)

☐ My building manager

☐ My supervisor

☐ The government

☐ The manufacturer of the devices

☐ My company

☐ Myself

☐ Other: ______________________

70. In the scenario, who do you think will benefit from having access to the collected data about you? (Choose all that apply)

☐ My building manager

☐ My supervisor

☐ The government

☐ The manufacturer of the devices

☐ My company

☐ Myself

☐ Other: ______________________

71. In the scenario, which of the following purposes do you think the collected data about you might be used for? (Choose all that apply)

☐ User profiling

☐ Localization

☐ Enforcing policies

☐ Improving smart devices

☐ Other: ____________________

☐ None

— Notification Preferences Questions —

Note that for post-scenario, we replace “at your workplace” with “in the scenario” to apply the context.

72. Do you want to be notified about the presence of the devices collecting data about you at your workplace?

☐ Yes

☐ No

73. Please briefly explain why you want (or why you don’t want) to be notified:

____

74. At your workplace, which of the following do you want to be notified about? (Choose all that apply)

☐ Presence of data collection (including types of data being collected)

☐ Purposes for which your data can be used

☐ How your data can be used

☐ For how long your data can be retained

☐ Who can access your data

☐ None of the above

75. How do you want to be notified? (Choose all that apply)

☐ Physical sign

☐ Mobile app

☐ Website

☐ Email

☐ Paper document

☐ Other: ____________________

— Demographic Information —

76. With which gender identity do you most identify?

☐ Male

☐ Female

☐ Other: ________________________

☐ Prefer not to answer

77. What is your age group?

☐ 18 - 24 years old

☐ 25 - 34 years old

☐ 35 - 44 years old

☐ 45 - 54 years old

☐ 55 - 64 years old

☐ 65 - 74 years old

☐ 75 years or older

☐ Prefer not to answer

78. What is the highest level of education you have completed?

☐ Some high school

☐ High school graduate

☐ Some college

☐ Associate’s degree (2-year college)

☐ Bachelor’s degree (4-year college)

☐ Graduate degree (Masters, PhD, MD, JD, etc.)

☐ Other: ________________________

☐ Prefer not to answer

79. What is your annual personal income before taxes (USD)?

☐ Less than $10,000

☐ $10,000 - $24,999

☐ $25,000 - $49,999

☐ $50,000 - $74,999

☐ $75,000 - $99,999

☐ $100,000 - $149,999

☐ $150,000 and greater

☐ Prefer not to answer

— Confidence and Understanding —

80. How confident are you with the Internet of Things technologies in general (e.g., voice assistants, smart security camera, drones, smart watches, virtual reality glass, etc.)?

☐ Extremely unconfident

☐ Somewhat unconfident

☐ Neither confident nor unconfident

☐ Somewhat confident

☐ Extremely confident

81. How would you describe your understanding of the Internet of Things?

☐ No understanding

☐ Below average

☐ Average

☐ Above average

☐ Strong understanding

Bibliography

[1] Lionel Sujay Vailshery. Number of internet of things (iot) connected devices worldwide from 2019 to 2021, with forecasts from 2022 to 2030, Nov 2022.

[2] Christopher Lentzsch, Sheel Jayesh Shah, Benjamin Andow, Martin Degeling, Anupam Das, and William Enck. Hey alexa, is this skill safe?: Taking a closer look at the alexa skill ecosystem. In 28th Annual Network and Distributed System Security Symposium (NDSS 2021). The Internet Society, 2021.

[3] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. In 2019 IEEE Symposium on Security and Privacy (SP), pages 1381–1396. IEEE, 2019.

[4] Jide Edu, Xavi Ferrer Aran, Jose Such, and Guillermo Suarez-Tangil. Skillvet: Automated traceability analysis of amazon alexa skills. IEEE Transactions on Dependable and Secure Computing, 2021.

[5] Jingjing Ren, Daniel J. Dubois, David Choffnes, Anna Maria Mandalari, Roman Kolcun, and Hamed Haddadi. Information exposure from consumer iot devices: A multidimensional, network-informed measurement approach. In Proceedings of the Internet Measurement Conference, IMC '19, page 267–279, New York, NY, USA, 2019. Association for Computing Machinery.

[6] Tamara Denning, Tadayoshi Kohno, and Henry M Levy. Computer security and the modern home. Communications of the ACM, 56(1):94–103, 2013.

[7] Qi Wang, Pubali Datta, Wei Yang, Si Liu, Adam Bates, and Carl A Gunter. Charting the attack surface of trigger-action iot platforms. In Proceedings of the 2019 ACM SIGSAC conference on computer and communications security, pages 1439–1453, 2019.

[8] Earlence Fernandes, Jaeyeon Jung, and Atul Prakash. Security analysis of emerging smart home applications. In 2016 IEEE symposium on security and privacy (SP), pages 636–654. IEEE, 2016.

[9] Xuejing Yuan, Yuxuan Chen, Aohui Wang, Kai Chen, Shengzhi Zhang, Heqing Huang, and Ian M Molloy. All your alexa are belong to us: A remote voice control attack against echo. In 2018 IEEE Global Communications Conference (GLOBECOM), pages 1–6. IEEE, 2018.

[10] Yuxuan Chen, Xuejing Yuan, Jiangshan Zhang, Yue Zhao, Shengzhi Zhang, Kai Chen, and XiaoFeng Wang. Devil's whisper: A general approach for physical adversarial attacks against commercial black-box speech recognition devices. In 29th {USENIX} Security Symposium ({USENIX} Security 20), pages 2667–2684, 2020.

[11] Tao Chen, Longfei Shangguan, Zhenjiang Li, and Kyle Jamieson. Metamorph: Injecting inaudible commands into over-the-air voice controlled systems. In Network and Distributed Systems Security (NDSS) Symposium, 2020.

[12] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. Hidden voice commands. In 25th USENIX Security Symposium (USENIX Security 16), pages 513–530, Austin, TX, 2016. USENIX Association.

[13] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey. Emerging threats in internet of things voice services. IEEE Security Privacy, 17(4):18–24, July 2019.

[14] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. Skill squatting attacks on amazon alexa. In 27th USENIX Security Symposium (USENIX Security 18), pages 33–47, Baltimore, MD, 2018. USENIX Association.

[15] Muhammad Ejaz Ahmed, Il-Youp Kwak, Jun Ho Huh, Iljoo Kim, Taekkyung Oh, and Hyoungshick Kim. Void: A fast and light voice liveness detection system. In 29th {USENIX} Security Symposium ({USENIX} Security 20), pages 2685–2702, 2020.

[16] Ajaya Neupane, Nitesh Saxena, Leanne M Hirshfield, and Sarah E Bratt. The crux of voice (in) security: A brain study of speaker legitimacy detection. In NDSS, 2019.

[17] Song Liao, Christin Wilson, Long Cheng, Hongxin Hu, and Huixing Deng. Measuring the effectiveness of privacy policies for voice assistant applications. In Annual Computer Security Applications Conference, ACSAC '20, page 856–869, New York, NY, USA, 2020. Association for Computing Machinery.

[18] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. Dangerous skills got certified: Measuring the trustworthiness of skill certification in voice personal assistant platforms. In Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, pages 1699–1716, 2020.

[19] Dawei Wang, Kai Chen, and Wei Wang. Demystifying the vetting process of voice-controlled skills on markets. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(3):1–28, 2021.

[20] David Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. Alexa, who am i speaking to?: Understanding users' ability to identify third-party apps on amazon alexa. ACM Transactions on Internet Technology (TOIT), 22(1):1–22, 2021.

[21] Noura Abdi, Kopo M Ramokapane, and Jose M Such. More than smart speakers: security and privacy perceptions of smart home personal assistants. In Fifteenth Symposium on Usable Privacy and Security ({SOUPS} 2019), pages 451–466, 2019.

[22] Nathaniel Fruchter and Ilaria Liccardi. Consumer attitudes towards privacy and security in home assistants. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, pages 1–6, 2018.

[23] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. Alexa, are you listening? privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–31, 2018.

[24] Nicole Meng, Dilara Keküllüoğlu, and Kami Vaniea. Owning and sharing: Privacy perceptions of smart speaker users. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–29, 2021.

[25] Noura Abdi, Xiao Zhan, Kopo M Ramokapane, and Jose Such. Privacy norms for smart home personal assistants. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2021.

[26] Dominik Pins, Timo Jakobi, Alexander Boden, Fatemeh Alizadeh, and Volker Wulf. Alexa, we need to talk: A data literacy approach on voice assistants. In Designing Interactive Systems Conference 2021, pages 495–507, 2021.

[27] Pardis Emami Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. Privacy expectations and preferences in an iot world. In Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), pages 399–412, Santa Clara, CA, July 2017. USENIX Association.

[28] Shikun Zhang, Yuanyuan Feng, Lujo Bauer, Lorrie Faith Cranor, Anupam Das, and Norman Sadeh. “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):282–304, 2021.

[29] Shikun Zhang, Yuanyuan Feng, and Norman Sadeh. Facial recognition: Understanding privacy concerns and attitudes across increasingly diverse deployment scenarios. In Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021), pages 243–262, 2021.

[30] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. Are those steps worth your privacy? fitness-tracker users’ perceptions of privacy and utility. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 5(4), dec 2022.

[31] Pardis Emami Naeini, Martin Degeling, Lujo Bauer, Richard Chow, Lorrie Faith Cranor, Mohammad Reza Haghighat, and Heather Patterson. The influence of friends and experts on privacy decision making in iot scenarios. Proc. ACM Hum.-Comput. Interact., 2(CSCW), November 2018.

[32] Natã Miccael Barbosa, Joon Sung Park, Yaxing Yao, and Yang Wang. “what if?” predicting individual users’ smart home privacy preferences and their changes. Proceedings on Privacy Enhancing Technologies, 2019:211 – 231, 2019.

[33] Noah Apthorpe, Yan Shvartzshnaider, Arunesh Mathur, Dillon Reisman, and Nick Feamster. Discovering smart home internet of things privacy norms using contextual integrity. 2(2), jul 2018.

[34] Leonardo Babun, Z. Berkay Celik, Patrick McDaniel, and A. Selcuk Uluagac. Real-time analysis of privacy-(un)aware iot applications. Proceedings on Privacy Enhancing Technologies, 2021(1):145–166, 2021.

[35] Amit Kumar Sikder, Leonardo Babun, Z. Berkay Celik, Abbas Acar, Hidayet Aksu, Patrick Mcdaniel, Engin Kirda, and Arif Selcuk Uluagac. Kratos: multi-user multi-device-aware access control system for the smart home. Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks, 2020.

[36] Council of European Union. General data protection regulation. https://gdpr-infor.eu, 2016.

[37] Office of the California Attorney General. California consumer privacy act (ccpa): First modified regulations. https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-text-of-mod-clean-020720.pdf, 2020.

[38] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. J. on Telecomm. & High Tech. L., 10:273, 2012.

[39] Anne Oeldorf-Hirsch and Jonathan A Obar. Overwhelming, important, irrelevant: Terms of service and privacy policy reading among older adults. In Proceedings of the 10th International Conference on Social Media and Society, pages 166–173, 2019.

[40] Nico Ebert, Kurt Alexander Ackermann, and Björn Scheppler. Bolder is better: Raising user awareness through salient and concise privacy notices. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–12, 2021.

[41] Pardis Emami-Naeini, Yuvraj Agarwal, Lorrie Faith Cranor, and Hanan Hibshi. Ask the experts: What should be on an iot privacy and security label? In 2020 IEEE Symposium on Security and Privacy (SP), pages 447–464. IEEE, 2020.

[42] Florian Schaub, Rebecca Balebako, Adam L Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In Eleventh Symposium On Usable Privacy and Security ({SOUPS} 2015), pages 1–17, 2015.

[43] CookieYes. Gdpr cookie consent banner examples, November 2019.

[44] Danny Yuxing Huang, Noah Apthorpe, Frank Li, Gunes Acar, and Nick Feamster. Iot inspector: Crowdsourcing labeled network traffic from smart home devices at scale. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 4(2), jun 2020.

[45] Hanbyul Choi, Jonghwa Park, and Yoonhyuk Jung. The role of privacy fatigue in online privacy behavior. Computers in Human Behavior, 81:42–51, 2018.

[46] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. IEEE Pervasive Computing, 17(3):35–46, 2018.

[47] Yuanyuan Feng, Yaxing Yao, and Norman Sadeh. A design space for privacy choices: Towards meaningful privacy control in the internet of things. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–16, 2021.

[48] Kristina Irion, Svetlana Yakovleva, Joris van Hoboken, M Thomson, et al. A roadmap to enhancing user control via privacy dashboards. 2017.

[49] Florian M. Farke, David G. Balash, Maximilian Golla, Markus Dürmuth, and Adam J. Aviv. Are privacy dashboards good for end users? evaluating user perceptions and reactions to google's my activity. In 30th USENIX Security Symposium (USENIX Security 21), pages 483–500. USENIX Association, August 2021.

[50] Philip Raschke, Axel Küpper, Olha Drozd, and Sabrina Kirrane. Designing a gdpr-compliant and usable privacy dashboard. In IFIP international summer school on privacy and identity management, pages 221–236. Springer, 2017.

[51] Denis Feth and Hartmut Schmitt. Requirement and quality models for privacy dashboards. In 2020 IEEE 7th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE), pages 1–6, 2020.

[52] Vandit Sharma and Mainack Mondal. Understanding and improving usability of data dashboards for simplified privacy control of voice assistant data. In 31st USENIX Security Symposium (USENIX Security 22), pages 3379–3395, Boston, MA, August 2022. USENIX Association.

[53] Sean Creamer. Kids Are Spending More Time with Voice, but Brands Shouldn't Rush to Engage Them, 2018.

[54] X. Yuan, Y. Chen, A. Wang, K. Chen, S. Zhang, H. Huang, and I. M. Molloy. All your alexa are belong to us: A remote voice control attack against echo. In 2018 IEEE Global Communications Conference (GLOBECOM), pages 1–6, Dec 2018.

[55] Reddit. Alexa is adding freetime skills that i cannot remove. https://www.reddit.com/r/alexa/comments/aba5u6/alexa_is_adding_freetime_skills_that_i_cannot/, 2018. Accessed: 2022-03-18.

[56] Amazon Forum. Can't remove/edit freetime content! anyone have a fix? https://www.amazonforum.com/forums/devices/fire-tablets/1815-cant-remove-edit-freetime-content-anyone-have-a, 2017. Accessed: 2022-03-18.

[57] Reddit. How can i know what the freetime unlimited skills are? https://www.reddit.com/r/amazonecho/comments/9nzuwj/how_can_i_know_what_the_freetime_unlimited_skills/, 2018. Accessed: 2022-03-18.

[58] Federal Trade Commission. Complying with coppa: Frequently asked questions. https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions, 2020. Accessed: 2022-03-18.

[59] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. “won’t somebody think of the children?” examining coppa compliance at scale. Proceedings on Privacy Enhancing Technologies, 2018(3):63–83, 2018.

[60] Emma Sorbring. Parents’ concerns about their teenage children’s internet use. Journal of Family Issues, 35(1):75–96, 2014.

[61] Larry D Rosen, Nancy A Cheever, and L Mark Carrier. The association of parenting style and child age with parental limit setting and adolescent myspace behavior. Journal of Applied Developmental Psychology, 29(6):459–471, 2008.

[62] Amazon. Alexa skills kit. https://developer.amazon.com/alexa/alexa-skills-kit, 2022. Accessed: 2022-03-18.

[63] Irwin Reyes, Primal Wijesekera, Abbas Razaghpanah, Joel Reardon, Narseo Vallina-Rodriguez, Serge Egelman, Christian Kreibich, et al. ” is our children’s apps learning?” automatically detecting coppa violations. In Workshop on Technology and Consumer Protection (ConPro 2017), in conjunction with the 38th IEEE Symposium on Security and Privacy (IEEE S&P 2017), 2017.

[64] Irwin Reyes, Primal Wijesekera, Joel Reardon, Amit Elazari Bar On, Abbas Razaghpanah, Narseo Vallina-Rodriguez, and Serge Egelman. ”won’t somebody think of the children?” examining coppa compliance at scale. PoPETs, 2018:63–83, 2018.

[65] Abbas Razaghpanah, Rishab Nithyanand, Narseo Vallina-Rodriguez, Srikanth Sundaresan, Mark Allman, Christian Kreibich, Phillipa Gill, et al. Apps, trackers, privacy, and regulators: A global study of the mobile tracking ecosystem. In The 25th Annual Network and Distributed System Security Symposium (NDSS 2018), 2018.

[66] Hamza Harkous, Kassem Fawaz, Kang G Shin, and Karl Aberer. Pribots: Conversational privacy with chatbots. In Twelfth Symposium on Usable Privacy and Security (SOUPS 2016), 2016.

[67] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems. 2019 IEEE Symposium on Security and Privacy (SP), pages 1381–1396, 2019.

[68] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. Dangerous Skills Got Certified: Measuring the Trustworthiness of Skill Certification in Voice Personal Assistant Platforms, page 1699–1716. Association for Computing Machinery, New York, NY, USA, 2020.

[69] Zhixiu Guo, Zijin Lin, Pan Li, and Kai Chen. SkillExplorer: Understanding the behavior of skills in large scale. In 29th USENIX Security Symposium (USENIX Security 20), pages 2649–2666. USENIX Association, August 2020.

[70] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven Bellovin, and Joel Reidenberg. Automated analysis of privacy requirements for mobile apps. In 2016 AAAI Fall Symposium Series, 2016.

[71] Natalija Vlajic, Marmara El Masri, Gianluigi M Riva, Marguerite Barry, and Derek Doran. Online tracking of kids and teens by means of invisible images: Coppa vs. gdpr. In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, pages 96–103. ACM, 2018.

[72] Joseph Turow. Privacy policies on children’s websites: Do they play by the rules? Report Series No. 33 at The Annenberg Public Policy Center of the University of Pennsylvania, 2001.

[73] Xiaomei Cai and Xiaoquan Zhao. Online advertising on popular children’s websites: Structural features and privacy issues. Computers in Human Behavior, 29(4):1510–1518, 2013.

[74] Gordon Chu, Noah Apthorpe, and Nick Feamster. Security and privacy analyses of internet of things children's toys. IEEE Internet of Things Journal, 6(1):978–985, 2018.

[75] Moustafa Mahmoud, Md Zakir Hossen, Hesham Barakat, Mohammad Mannan, and Amr Youssef. Towards a comprehensive analytical framework for smart toy privacy practices. In Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust, pages 64–75. ACM, 2018.

[76] Joshua Streiff, Olivia Kenny, Sanchari Das, Andrew Leeth, and L Jean Camp. Who's watching your child? exploring home security risks with smart toy bears. In 2018 IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation (IoTDI), pages 285–286. IEEE, 2018.

[77] Junia Valente and Alvaro A Cardenas. Security & privacy in smart toys. In Proceedings of the 2017 Workshop on Internet of Things Security and Privacy, pages 19–24. ACM, 2017.

[78] Laura Rafferty, Patrick CK Hung, Marcelo Fantinato, Sarajane Marques Peres, Farkhund Iqbal, Sy-Yen Kuo, and Shih-Chia Huang. Towards a privacy rule conceptual model for smart toys. In Computing in Smart Toys, pages 85–102. Springer, 2017.

[79] Jeffrey Haynes, Maribette Ramirez, Thaier Hayajneh, and Md Zakirul Alam Bhuiyan. A framework for preventing the exploitation of iot smart toys for reconnaissance and exfiltration. In International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, pages 581–592. Springer, 2017.

[80] Pekka Mertala. Young children's perceptions of ubiquitous computing and the internet of things. British Journal of Educational Technology, 51(1):84–102, 2020.

[81] Andrew Manches, Pauline Duncan, Lydia Plowman, and Shari Sabeti. Three questions about the internet of things and children. TechTrends, 59(1):76–83, 2015.

[82] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. Toys that listen: A study of parents, children, and internet-connected toys. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pages 5197–5207. ACM, 2017.

[83] Jun Zhao, Ge Wang, Carys Dally, Petr Slovak, Julian Edbrooke-Childs, Max Van Kleek, and Nigel Shadbolt. 'i make up a silly name': Understanding children's perception of privacy risks online. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, pages 106:1–106:13, New York, NY, USA, 2019. ACM.

[84] Priya Kumar, Shalmali Milind Naik, Utkarsha Ramesh Devkar, Marshini Chetty, Tamara L Clegg, and Jessica Vitak. 'no telling passcodes out because they're private': Understanding children's mental models of privacy and security online. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW):64, 2017.

[85] Tehila Minkus, Kelvin Liu, and Keith W Ross. Children seen but not heard: When parents compromise children's online privacy. In Proceedings of the 24th International Conference on World Wide Web, pages 776–786. International World Wide Web Conferences Steering Committee, 2015.

[86] Eszter Hargittai, Jason Schultz, John Palfrey, et al. Why parents help their children lie to facebook about age: Unintended consequences of the 'children's online privacy protection act'. First Monday, 16(11), 2011.

[87] Yangyong Zhang, Lei Xu, Abner Mendoza, Guangliang Yang, Phakpoom Chinprutthiwong, and Guofei Gu. Life after speech recognition: Fuzzing semantic misinterpretation for voice assistant applications. In Proc. of the Network and Distributed System Security Symposium (NDSS'19), 2019.

[88] Amazon. Understand name-free interactions. https://developer.amazon.com/docs/custom-skills/understand-name-free-interaction-for-custom-skills.html, 2022. Accessed: 2022-03-18.

[89] Amazon. Amazon alexa skills. https://www.amazon.com/alexa-skills/b?ie=UTF8&node=13727921011, 2022. Accessed: 2022-03-18.

[90] Leonard Richardson. Beautifulsoup. https://www.crummy.com/software/BeautifulSoup/, 2021. Accessed: 2022-03-18.

[91] Amazon. Alexa simulator. https://developer.amazon.com/docs/devconsole/test-your-skill.html#test-simulator, 2022. Accessed: 2022-03-18.

[92] Amazon. Understand how users invoke custom skills. https://developer.amazon.com/docs/custom-skills/understanding-how-users-invoke-custom-skills.html, 2022. Accessed: 2022-03-18.

[93] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. spaCy: Industrial-strength Natural Language Processing in Python, 2020.

[94] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In Association for Computational Linguistics (ACL) System Demonstrations, pages 55–60, 2014.

[95] Peng Qi, Timothy Dozat, Yuhao Zhang, and Christopher D. Manning. Universal dependency parsing from scratch. In Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pages 160–170, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[96] Ann Taylor, Mitchell Marcus, and Beatrice Santorini. The penn treebank: an overview. In Treebanks, pages 5–22. Springer, 2003.

[97] Universal Dependencies contributors. Universal pos tags. https://universaldependencies.org/u/pos/, 2021. Accessed: 2022-03-18.

[98] Harish Tayyar Madabushi and Mark Lee. High accuracy rule-based question classification using question syntax and semantics. In Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pages 1220–1230, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

[99] Xin Li and Dan Roth. Definition of question classes. https://cogcomp.seas.upenn.edu/Data/QA/QC/definition.html, 2002. Accessed: 2022-03-18.

[100] Amazon. Alexa simulator limitations. https://developer.amazon.com/docs/devconsole/test-your-skill.html#alexa-simulator-limitations, 2022. Accessed: 2022-03-18.

[101] WebPurify. Webpurify for children's apps and websites. https://www.webpurify.com/childrens-apps-websites/, 2022. Accessed: 2022-03-18.

[102] Microsoft. Azure content moderator. https://docs.microsoft.com/en-us/azure/cognitive-services/content-moderator/, 2022. Accessed: 2022-03-18.

[103] DoD Privacy Office. Frequently answered questions. https://dpcld.defense.gov/Privacy/About-the-Office/FAQs/#2, 2022. Accessed: 2022-03-18.

[104] Robert J. Fisher and James E. Katz. Social-desirability bias and the validity of self-reported values. Psychology & Marketing, 17(2):105–120, 2000.

[105] Grand View Research. Intelligent virtual assistant market size report. https://developer.amazon.com/en-US/docs/alexa/custom-skills/configure-permissions-for-customer-information-in-your-skill.html, 2020.

[106] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. Skill squatting attacks on amazon alexa. In 27th {USENIX} Security Symposium ({USENIX} Security 18), pages 33–47, 2018.

[107] Mary K Bispham, Ioannis Agrafiotis, and Michael Goldsmith. Nonsense attacks on google assistant and missense attacks on amazon alexa. 2019.

[108] Takeshi Sugawara, Benjamin Cyr, Sara Rampazzi, Daniel Genkin, and Kevin Fu. Light commands: laser-based audio injection attacks on voice-controllable systems. In 29th {USENIX} Security Symposium ({USENIX} Security 20), pages 2631–2648, 2020.

[109] Fabian Braunlein and Luise Frerichs. Smart spies: Alexa and google home expose users to vishing and eavesdropping. https://www.srlabs.de/bites/smart-spies, 2019.

[110] Yangyong Zhang, Lei Xu, Abner Mendoza, Guangliang Yang, Phakpoom Chinprutthiwong, and Guofei Gu. Life after speech recognition: Fuzzing semantic misinterpretation for voice assistant applications. In Proc. of the Network and Distributed System Security Symposium (NDSS'19), 2019.

[111] Jeffrey Young, Song Liao, Long Cheng, Hongxin Hu, and Huixing Deng. SkillDetective: Automated Policy-Violation detection of voice assistant applications in the wild. In 31st USENIX Security Symposium (USENIX Security 22), pages 1113–1130, Boston, MA, August 2022. USENIX Association.

[112] Amazon. Configure permissions for customer information in your skill. https://developer.amazon.com/en-US/docs/alexa/custom-skills/configure-permissions-for-customer-information-in-your-skill.html, 2022.

[113] Allen Institute for AI. Textual entailment. https://demo.allennlp.org/textual-entailment/elmo-snli, 2022.

[114] Ankur P Parikh, Oscar Täckström, Dipanjan Das, and Jakob Uszkoreit. A decomposable attention model for natural language inference. arXiv preprint arXiv:1606.01933, 2016.

[115] Lutz Prechelt. Early Stopping — But When?, pages 53–67. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[116] Amazon. What is the alexa voice service? https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/get-started-with-alexa-voice-service.html, 2022.

[117] Amazon. Set up the avs device sdk on ubuntu. https://developer.amazon.com/en-US/docs/alexa/avs-device-sdk/ubuntu.html, 2022.

[118] Tu Le, Danny Yuxing Huang, Noah Apthorpe, and Yuan Tian. Skillbot: Identifying risky content for children in alexa skills. ACM Trans. Internet Technol., 22(3), jul 2022.

[119] Amazon. Amazon's alexa tells 10-year-old child to touch penny to exposed plug socket. https://www.cnn.com/2021/12/29/business/amazon-alexa-penny-plug-intl-scli/index.html, 2022. Accessed: 2022-03-18.

[120] Oliver Budzinski, Victoriia Noskova, and Xijie Zhang. The brave new world of digital personal assistants: Benefits and challenges from an economic perspective. NETNOMICS: Economic Research and Electronic Networking, 20(2):177–194, 2019.

[121] Ramin Yaghoubzadeh, Marcel Kramer, Karola Pitsch, and Stefan Kopp. Virtual agents as daily assistants for elderly or cognitively impaired people. In International workshop on intelligent virtual agents, pages 79–91. Springer, 2013.

[122] Angela J. Campbell and Lindsey Barrett. In the matter of request for investigation of amazon, inc.'s echo dot kids edition for violating the children's online privacy protection act. Letter to Federal Trade Commission, May 2019.

[123] Daniel J Dubois, Roman Kolcun, Anna Maria Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. When speakers are all ears: Characterizing misactivations of iot smart speakers. Proceedings on Privacy Enhancing Technologies, 2020(4):255–276, 2020.

[124] Andrew McCarthy, Benedict R Gaster, and Phil Legg. Shouting through letterboxes: A study on attack susceptibility of voice assistants. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pages 1–8. IEEE, 2020.

[125] Tu Le, Danny Yuxing Huang, Noah Apthorpe, and Yuan Tian. Skillbot: Identifying risky content for children in alexa skills. ACM Trans. Internet Technol., 22(3), Jul 2022.

[126] Lea Schönherr, Maximilian Golla, Thorsten Eisenhofer, Jan Wiele, Dorothea Kolossa, and Thorsten Holz. Unacceptable, where is my privacy? exploring accidental triggers of smart speakers, 2020.

[127] Rebecca Adaimi, Howard Yong, and Edison Thomaz. Ok google, what am i doing? acoustic activity recognition bounded by conversational assistant interactions. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(1):1–24, 2021.

[128] Luigi De Russis and Alberto Monge Roffarello. On the benefit of adding user preferences to notification delivery. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, page 1561–1568, New York, NY, USA, 2017. Association for Computing Machinery.

[129] Eric Zeng and Franziska Roesner. Understanding and improving security and privacy in Multi-User smart homes: A design exploration and In-Home user study. In 28th USENIX Security Symposium (USENIX Security 19), pages 159–176, Santa Clara, CA, August 2019. USENIX Association.

[130] Alexandra Voit, Thomas Kosch, Henrike Weingartner, and Paweł W Woźniak. The attention kitchen: Comparing modalities for smart home notifications in a cooking scenario. In Proceedings of the 20th International Conference on Mobile and Ubiquitous Multimedia, pages 90–97, 2021.

[131] Parth Kirankumar Thakkar, Shijing He, Shiyu Xu, Danny Yuxing Huang, and Yaxing Yao. "it would probably turn into a social faux-pas": Users' and bystanders' preferences of privacy awareness mechanisms in smart homes. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery.

[132] T. Le, A. Wang, Y. Yao, Y. Feng, A. Heydarian, N. Sadeh, and Y. Tian. Exploring smart commercial building occupants' perceptions and notification preferences of internet of things data collection in the united states. In 2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P), pages 1030–1046, Los Alamitos, CA, USA, Jul 2023. IEEE Computer Society.

[133] Stanley Presser, Mick P Couper, Judith T Lessler, Elizabeth Martin, Jean Martin, Jennifer M Rothgeb, and Eleanor Singer. Methods for testing and evaluating survey questions. Methods for testing and evaluating survey questionnaires, pages 1–22, 2004.

[134] Andreas Wagner, William O'Brien, and Bing Dong. Exploring occupant behavior in buildings. Wagner, A., O'Brien, W., Dong, B., Eds, 2018.

[135] Joseph G Allen and John D Macomber. Healthy buildings: How indoor spaces drive performance and productivity. Harvard University Press, 2020.

[136] Jiakang Lu, Tamim Sookoor, Vijay Srinivasan, Ge Gao, Brian Holben, John Stankovic, Eric Field, and Kamin Whitehouse. The smart thermostat: using occupancy sensors to save energy in homes. In Proceedings of the 8th ACM conference on embedded networked sensor systems, pages 211–224, 2010.

[137] MG Figueiro, B Steverson, J Heerwagen, R Yucel, C Roohan, L Sahin, K Kampschroer, and MS Rea. Light, entrainment and alertness: A case study in offices. Lighting Research & Technology, 52(6):736–750, 2020.

[138] Heather Richter Lipford, Madiha Tabassum, Paritosh Bahirat, Yaxing Yao, and Bart P Knijnenburg. Privacy and the internet of things. Modern Socio-Technical Perspectives on Privacy, page 233, 2022.

[139] Stefany Cruz, Logan Danek, Shinan Liu, Christopher Kraemer, Zixin Wang, Nick Feamster, Danny Yuxing Huang, Yaxing Yao, and Josiah Hester. Augmented reality's potential for identifying and mitigating home privacy leaks. arXiv preprint arXiv:2301.11998, 2023.

[140] Scott Harper, Maryam Mehrnezhad, and John Mace. User privacy concerns in commercial smart buildings. Journal of Computer Security, (Preprint):1–33, 2022.

[141] Alexandra Voit, Dominik Weber, Yomna Abdelrahman, Marie Salm, Paweł W. Woźniak, Katrin Wolf, Stefan Schneegass, and Niels Henze. Exploring non-urgent smart home notifications using a smart plant system. In 19th International Conference on Mobile and Ubiquitous Multimedia, MUM '20, page 47–58, New York, NY, USA, 2020. Association for Computing Machinery.

[142] Primal Pappachan, Martin Degeling, Roberto Yus, Anupam Das, Sruti Bhagavatula, William Melicher, Pardis Emami Naeini, Shikun Zhang, Lujo Bauer, Alfred Kobsa, Sharad Mehrotra, Norman Sadeh, and Nalini Venkatasubramanian. Towards privacy-aware smart buildings: Capturing, communicating, and enforcing privacy policies and preferences. In 2017 IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW), pages 193–198, 2017.

[143] Haojian Jin, Boyuan Guo, Rituparna Roychoudhury, Yaxing Yao, Swarun Kumar, Yuvraj Agarwal, and Jason I Hong. Exploring the needs of users for supporting privacy-protective behaviors in smart homes. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pages 1–19, 2022.

[144] Michelle Goddard. The eu general data protection regulation (gdpr): European regulation that has a global impact. International Journal of Market Research, 59(6):703–705, 2017.

[145] Eric Goldman. An introduction to the california consumer privacy act (ccpa). Santa Clara Univ. Legal Studies Research Paper, 2020.

[146] Matthias Caesar and Jan Steffan. A location privacy analysis of bluetooth mesh. Journal of Information Security and Applications, 54:102563, 2020.

[147] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. The visual microphone: Passive recovery of sound from video. 2014.

[148] Michael Rubinstein et al. Analysis and visualization of temporal variations in video. PhD thesis, Massachusetts Institute of Technology, 2014.

[149] Farrokh Jazizadeh, Burcin Becerik-Gerber, Mario Berges, and Lucio Soibelman. An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems. Advanced Engineering Informatics, 28(4):311–326, 2014.

[150] Abbas Acar, Hossein Fereidooni, Tigist Abera, Amit Kumar Sikder, Markus Miettinen, Hidayet Aksu, Mauro Conti, Ahmad-Reza Sadeghi, and Selcuk Uluagac. Peek-a-boo: I see your smart home activities, even encrypted! In Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks, WiSec '20, page 207–218, New York, NY, USA, 2020. Association for Computing Machinery.

[151] Luis Puche Rondon, Leonardo Babun, Ahmet Aris, Kemal Akkaya, and Arif Selcuk Uluagac. Poisonivy: (in)secure practices of enterprise iot systems in smart buildings. Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2020.

[152] Luis Puche Rondon, Leonardo Babun, Ahmet Aris, Kemal Akkaya, and A. Selcuk Uluagac. Survey on enterprise internet-of-things systems (e-iot): A security perspective, 2021.

[153] A survey on iot platforms: Communication, security, and privacy perspectives. Computer Networks, 192, June 2021.

[154] Yaxing Yao, Justin Reed Basdeo, Smirity Kaushik, and Yang Wang. Defending my castle: A co-design study of privacy mechanisms for smart homes. In Proceedings of the 2019 chi conference on human factors in computing systems, pages 1–12, 2019.

[155] Yaxing Yao, Justin Reed Basdeo, Oriana Rosata Mcdonough, and Yang Wang. Privacy perceptions and designs of bystanders in smart homes. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–24, 2019.

[156] Stephan Cejka, Felix Knorr, and Florian Kintzler. Privacy issues in smart buildings by examples in smart metering. 2019.

[157] Tong Wu, Murtadha Aldeer, Tahiya Chowdhury, Amber Haynes, Fateme Nikseresht, Mahsa Pahlavikhah Varnosfaderani, Jiechao Gao, Arsalan Heydarian, Brad Campbell, and Jorge Ortiz. The smart building privacy challenge. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, pages 238–239, 2021.

[158] Stanley Presser, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, Jennifer M. Rothgeb, and Eleanor Singer. Methods for Testing and Evaluating Survey Questions, chapter 1, pages 1–22. John Wiley & Sons, Ltd, 2004.

[159] Parth Kirankumar Thakkar, Shijing He, Shiyu Xu, Danny Yuxing Huang, and Yaxing Yao. "it would probably turn into a social faux-pas": Users' and bystanders' preferences of privacy awareness mechanisms in smart homes. In CHI Conference on Human Factors in Computing Systems, pages 1–13, 2022.

[160] Jennifer Huddleston. The price of privacy: The impact of strict data regulations on innovation and more. https://www.americanactionforum.org/insight/the-price-of-privacy-the-impact-of-strict-data-regulations-on-innovation-and-more/, 2021.

[161] NPR. The Smart Audio Report, 2019.

[162] Amazon. What are upcs, eans, isbns. and asins? https://www.amazon.com/gp/seller/asin-upc-isbn-info.html, 2022. Accessed: 2022-03-18.

[163] Mozilla. How connected are you? take mozilla's survey. https://medium.com/mozilla-internet-citizen/how-connected-are-you-take-mozillas-survey-89c6c0ded21, 2017. Accessed: 2022-03-18.

[164] Sarah Perez. Amazon adds parental consent to alexa skills aimed at children, launches first legal kids' skills. https://techcrunch.com/2017/08/31/amazon-adds-parental-consent-to-alexa-skills-aimed-at-children-launches-first-legal-k 2017. Accessed: 2022-03-18.

[165] Amazon. Revise and update your skill after publication. https://developer.amazon.com/docs/devconsole/test-and-submit-your-skill.html#revise-and-update, 2022. Accessed: 2022-03-18.

[166] Amazon. Policy testing for an alexa skill. https://developer.amazon.com/en-US/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html#cert-content-ratings, 2022. Accessed: 2022-03-18.

[167] Eric J. Johnson and Daniel Goldstein. Do defaults save lives? Science, 302(5649):1338–1339, 2003.

[168] Brigitte C Madrian and Dennis F Shea. The power of suggestion: Inertia in 401 (k) participation and savings behavior. The Quarterly journal of economics, 116(4):1149–1187, 2001.

[169] Craig RM McKenzie, Michael J Liersch, and Stacey R Finkelstein. Recommendations implicit in policy defaults. Psychological Science, 17(5):414–420, 2006.

[170] Peter V Marsden and James D Wright. Handbook of survey research. Emerald Group Publishing, 2010.

[171] Amazon. Host a custom skill as a web service. https://developer.amazon.com/docs/custom-skills/host-a-custom-skill-as-a-web-service.html, 2022. Accessed: 2022-03-18.

[172] Plex Forum. Alexa skill not kid friendly (freetime)? https://forums.plex.tv/t/alexa-skill-not-kid-friendly-freetime/343477, 2018. Accessed: 2022-03-18.

[173] Amazon Forum. Freetime unlimited alexa skills not available in parent dashboard. https://www.amazonforum.com/forums/devices/echo-alexa/497656-freetime-unlimited-alexa-skills-not-available-in, 2018. Accessed: 2022-03-18.

[174] Statista. Total number of amazon alexa skills in selected countries. https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/, 2022. Accessed: 2022-03-18.

[175] Amazon. Interaction model schemas. https://developer.amazon.com/docs/smapi/interaction-model-schema.html#interaction-model-schemas, 2022. Accessed: 2022-03-18.

[176] Logan Blue, Hadi Abdullah, Luis Vargas, and Patrick Traynor. 2ma: Verifying voice commands via two microphone authentication. In Proceedings of the 2018 on Asia Conference on Computer and Communications Security, pages 89–100. ACM, 2018.

[177] Amr Alanwar, Bharathan Balaji, Yuan Tian, Shuo Yang, and Mani Srivastava. Echosafe: Sonar-based verifiable interaction with intelligent digital agents. In Proceedings of the 1st ACM Workshop on the Internet of Safe Things, pages 38–43. ACM, 2017.

[178] Logan Blue, Luis Vargas, and Patrick Traynor. Hello, is it me you're looking for?: Differentiating between human and electronic speakers for voice interface security. In Proceedings of the 11th ACM Conference on Security & Privacy in Wireless and Mobile Networks, pages 123–133. ACM, 2018.

[179] Erkam Uzun, Simon Pak Ho Chung, Irfan Essa, and Wenke Lee. rtcaptcha: A real-time captcha based liveness detection system. 2018.

[180] Linghan Zhang, Sheng Tan, and Jie Yang. Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pages 57–71. ACM, 2017.

[181] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. Cocaine noodles: Exploiting the gap between human and machine speech recognition. In 9th USENIX Workshop on Offensive Technologies (WOOT 15), Washington, D.C., 2015. USENIX Association.

[182] Nirupam Roy, Haitham Hassanieh, and Romit Roy Choudhury. Backdoor: Making microphones hear inaudible sounds. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pages 2–14. ACM, 2017.

[183] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. Inaudible voice commands: The long-range attack and defense. In 15th USENIX Symposium on Networked Systems Design and Implementation (NSDI 18), pages 547–560, Renton, WA, 2018. USENIX Association.

[184] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. Dolphinattack: Inaudible voice commands. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS '17, pages 103–117, New York, NY, USA, 2017. ACM.

[185] Yeongjin Jang, Chengyu Song, Simon P. Chung, Tielei Wang, and Wenke Lee. A11y attacks: Exploiting accessibility in operating systems. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, CCS '14, pages 103–115, New York, NY, USA, 2014. ACM.

[186] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices, SPSM '14, pages 63–74, New York, NY, USA, 2014. ACM.

[187] US Census Bureau. Pinc-01. selected characteristics of people 15 years and over, by total money income, work experience, race, hispanic origin, and sex. https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-01.html, 2021.

[188] US Census Bureau. 2017 american community survey. https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2017/, 2017.

[189] Jacob Cohen. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1):37–46, 1960.

[190] Juliet Corbin and Anselm Strauss. Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications, 2014.

[191] Microsoft. Azure content moderator's text moderation concepts. https://docs.microsoft.com/en-us/azure/cognitive-services/content-moderator/text-moderation-api, 2022. Accessed: 2022-03-18.

[192] Microsoft. Azure content moderator client library for python. https://docs.microsoft.com/en-us/azure/cognitive-services/content-moderator/python-sdk-quickstart, 2022. Accessed: 2022-03-18.

[193] WebPurify. Webpurify api. https://www.webpurify.com/documentation/, 2022. Accessed: 2022-03-18.

[194] Eric Zeng and Franziska Roesner. Understanding and improving security and privacy in multi-user smart homes: A design exploration and in-home user study. In 28th USENIX Security Symposium (USENIX Security 19), pages 159–176, Santa Clara, CA, August 2019. USENIX Association.

[195] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. Statistical methods for rates and proportions. john wiley & sons, 2013.

[196] Z Berkay Celik, Gang Tan, and Patrick D McDaniel. Iotguard: Dynamic enforcement of security and safety policy in commodity iot. In NDSS, 2019.

[197] Xuejing Yuan, Yuxuan Chen, Yue Zhao, Yunhui Long, Xiaokang Liu, Kai Chen, Shengzhi Zhang, Heqing Huang, Xiaofeng Wang, and Carl A Gunter. Commandersong: A systematic approach for practical adversarial voice recognition. In 27th {USENIX} Security Symposium ({USENIX} Security 18), pages 49–64, 2018.

[198] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. Rethinking access control and authentication for the home internet of things (iot). In 27th {USENIX} Security Symposium ({USENIX} Security 18), pages 255–272, 2018.

[199] Song Liao, Christin Wilson, Cheng Long, Hongxin Hu, and Huixing Deng. Problematic privacy policies of voice assistant applications. IEEE Security & Privacy, 2021.

[200] Faysal Hossain Shezan, Hang Hu, Jiamin Wang, Gang Wang, and Yuan Tian. Read between the lines: An empirical measurement of sensitive applications of voice personal assistant systems. In Proceedings of The Web Conference 2020, pages 1006–1017, 2020.

[201] Faysal Hossain Shezan, Hang Hu, Gang Wang, and Yuan Tian. Verhealth: Vetting medical voice applications through policy enforcement. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(4):1–21, 2020.

[202] Nora Ni Loideain and Rachel Adams. From alexa to siri and the gdpr: the gendering of virtual personal assistants and the role of data protection impact assessments. Computer Law & Security Review, 36:105366, 2020.

[203] Nathan L Tenhundfeld, Hannah M Barr, HO Emily, and Kristin Weger. Is my siri the same as your siri? an exploration of users' mental model of virtual personal assistants, implications for trust. IEEE Transactions on Human-Machine Systems, 2021.

[204] Kiljae Lee, Kyung Young Lee, and Lorn Sheehan. Hey alexa! a magic spell of social glue?: Sharing a smart voice assistant speaker and its impact on users' perception of group harmony. Information Systems Frontiers, 22(3):563–583, 2020.

[205] Clemens Krueger and Sean McKeown. Using amazon alexa apis as a source of digital evidence. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pages 1–8. IEEE, 2020.

[206] M Vimalkumar, Sujeet Kumar Sharma, Jang Bahadur Singh, and Yogesh K Dwivedi. 'okay google, what about my privacy?': User's privacy perceptions and acceptance of voice based digital assistants. Computers in Human Behavior, 120:106763, 2021.

[207] Katherine Taken Smith. Marketing via smart speakers: what should alexa say? Journal of Strategic Marketing, 28(4):350–365, 2020.

[208] Roei Schuster, Vitaly Shmatikov, and Eran Tromer. Situational access control in the internet of things. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, pages 1056–1073, 2018.

[209] Amazon. Alexa voice service. https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/get-started-with-alexa-voice-service.html, 2022. Accessed: 2022-01-30.

[210] E. Fernandes, J. Jung, and A. Prakash. Security analysis of emerging smart home applications. In 2016 IEEE Symposium on Security and Privacy (SP), pages 636–654, May 2016.

[211] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. Rethinking access control and authentication for the home internet of things (iot). In 27th USENIX Security Symposium (USENIX Security 18), pages 255–272, Baltimore, MD, 2018. USENIX Association.

[212] Pardis Emami Naeini, Martin Degeling, Lujo Bauer, Richard Chow, Lorrie Faith Cranor, Mohammad Reza Haghighat, and Heather Patterson. The influence of friends and experts on privacy decision making in iot scenarios. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–26, 2018.

[213] Natã Miccael Barbosa, Joon S Park, Yaxing Yao, and Yang Wang. " what if?" predicting individual users' smart home privacy preferences and their changes. PoPETs, 2019(4):211–231, 2019.

[214] Terry L Buchanan, Kenneth N Barker, J Tyrone Gibson, Bernard C Jiang, and Robert E Pearson. Illumination and errors in dispensing. American journal of hospital pharmacy, 48(10):2137–2145, 1991.