"Alexa, Is Dynamic Content Safe?" Understanding the Risks of Dynamic Content in the Alexa Skill Ecosystem

Nathan McClaran
nmcclaran@tamu.edu
Texas A&M University
College Station, Texas, United States

Payton Walker
paytonwalker@mitre.org
Texas A&M University
College Station, Texas, United States

Zihao Zheng
zzh523710043@tamu.edu
Texas A&M University
College Station, Texas, United States

Yangyong Zhang
yangyongzhang.io@gmail.com
Texas A&M University
College Station, Texas, United States

Nitesh Saxena
nsaxena@tamu.edu
Texas A&M University
College Station, Texas, United States

Guofei Gu
guofei@cse.tamu.edu
Texas A&M University
College Station, Texas, United States

Abstract

Despite the increasing popularity of voice assistants such as Amazon Alexa, the security implications of dynamic skill content (content modifiable without resubmission) in voice assistant skills (voice-activated applications) remain largely unexplored. This paper presents the first large-scale analysis of Alexa's dynamic content ecosystem using D-Explorer, a ChatGPT powered chatbot. From a dataset of 10,407 skill interactions, we investigate: 1) the mechanisms of Alexa dynamic content, 2) the associated security risks, and 3) the prevalence of these risks in published skills. Our analysis reveals that 34% of skills contain dynamic content in interactions, 95% access external resources (increasing attack vectors), 7% of skill conversations exhibit problematic (potentially harmful or privacy-infringing) interactions related to dynamic content, and 90% of skills connect to a potentially vulnerable dynamic resource during interaction. These findings expose significant vulnerabilities, highlighting the critical need for stricter developer rules and security measures to prevent unpredictable, harmful, and privacy compromising interactions within the Alexa skill ecosystem.

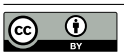
CCS Concepts

• **Security and privacy → Domain-specific security and privacy architectures**; *Information accountability and usage control*; Systems security.

Keywords

Amazon Alexa, Internet of Things Security, Dynamic Content

ACM Reference Format:
Nathan McClaran, Payton Walker, Zihao Zheng, Yangyong Zhang, Nitesh Saxena, and Guofei Gu. 2025. "Alexa, Is Dynamic Content Safe?" Understanding the Risks of Dynamic Content in the Alexa Skill Ecosystem. In *18th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec 2025), June 30-July 3, 2025, Arlington, VA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3734477.3734701



WiSec 2025, Arlington, VA, USA

ACM ISBN 979-8-4007-1530-3/2025/06
https://doi.org/10.1145/3734477.3734701

1 Introduction

The Internet of Things (IoT) is becoming a core part of society, with an estimated 18.8 billion devices connected by the end of 2024 [44]. One of the most common IoT devices is a smart speaker, with 35% adults in the United States owning one in 2022 [28]. Amazon Alexa, the most popular smart speaker [11], owes its success in part to its versatile skills, which range from smart home management to music and shopping [6]. These skills, akin to applications in Apple and Google's app stores [7, 26, 30], are deployed using Alexa's skill ecosystem, which supports both native and third-party skills [8]. The skill ecosystem has grown rapidly, increasing by 770% since its inception, with more than 100,000 skills available [33].

A crucial component of the skill ecosystem is dynamic content. Amazon defines dynamic content as content that pulls data from an external source as part of the functionality of the skill [31], and states that "dynamic content ensures that users will want to continue to interact with your Alexa skill" [31]. Dynamic content allows the generation of new or updated user content without code changes through external resource buckets or connections to websites and web resources. For skills reporting real-time data, such as weather, transportation, or financial information, dynamic content is essential. Based on Amazon's post on implementing dynamic content [31], **we define dynamic content as any resource that has the potential to change the responses of a skill without needing code changes or skill redeployment.**

While dynamic content offers benefits, it also poses security concerns, partially mirroring mobile app and web security challenges. Like those ecosystems, Alexa skills often use hidden resources, and many proposed attacks have roots in web/app security research. The Alexa skill ecosystem presents a unique attack surface. This is due to: skills and resources often hosted in the AWS cloud, leading to shared vulnerabilities; hidden skill endpoints that evade common traffic analysis; a lack of versioning, obscuring resource or behavior changes; and ambiguous voice commands which, unlike web/app interactions, lack user reviews, ratings, or descriptions, further obscuring skill interactions. Consequently, exploits using dynamic content can have a greater impact. Alexa users' high trust in Amazon, coupled with their unawareness of third-party data storage, creates vulnerability. As a whole-house device with shared account interactions, a single exploit can affect multiple Alexa devices and users. Furthermore, the opaque backend code makes skill auditing

nearly impossible for users, underscoring the need to monitor dynamic content to prevent exposure to inappropriate or dangerous material.

Despite recent work on skill-related security issues, such as those focusing on skill squatting and policy violations, Alexa skill dynamic content remains largely unexplored. To our knowledge, no previous studies have examined the usage and risks of *dynamic content* within the Alexa skill ecosystem.

1.1 Challenges

We encountered two primary challenges: detecting dynamic content and detecting external skill resources.

1.1.1 Detecting Dynamic Content. Although most Alexa skills display dynamic content warnings, these warnings do not consistently indicate actual dynamic behavior, suggesting that these warnings are often defaults. Simply detecting calls to external resources is insufficient to confirm dynamic content, as many such calls are not directly observable through traffic or data analysis. Cloud functions and remote servers, commonly used for dynamic resource calls, further complicate our analysis.

Solution: We identified dynamic content through conversational analysis, noting indicators like requests for updated schedules (ticket skills) or recent headlines (news skills). Many skills also disclose website resources in conversation, as per policy. By analyzing these conversational cues, we reliably identified dynamic content usage and any resulting problematic content. To automate this process, we developed a chatbot. We did not compare generated responses over time due to the unreliable nature of conversational interactions for detecting temporal changes. This solution allowed us to create a reliable method of detection, where no other method was available within our limitations.

1.1.2 Detecting Skill Resources. Despite understanding Alexa's dynamic content development, skill resources remain largely opaque, hindering security analysis. Many resources are inaccessible without successful skill interactions, and cloud-based resource calls evade local traffic analysis. Thus, we needed a method to locate resources without code access, as they are only available during skill execution.

Solution: We found that skill resources are often directly retrieved remotely. By inspecting the Alexa developer portal's page source during skill execution (automated via our chatbot), we extracted external resource URLs. These links were only accessible during successful skill invocation and interaction, allowing us to uncover otherwise hidden resources.

1.2 Findings

Our large-scale security analysis revealed that of the top 20% most popular Alexa skills a significant portion contain potential security and privacy risks. 34% of skills contain dynamic content in interactions, potentially allowing harmful or malicious content and 95% access external resources, broadening the attack surface. Alarmingly 7% of skill conversations exhibit problematic interactions (e.g. inappropriate content, privacy violations) related to dynamic content. A particularly concerning finding is the higher prevalence of problematic content in child-focused categories versus other categories: of the skills found with profanity or inappropriate content, 91% of total occurrences and 77% of inappropriate content occurrences were found in skill categories commonly used by children (Education & Reference, Kids, and Novelty & Humor). Furthermore, many skill resources exhibit technical vulnerabilities: 90% of the skills connected to a vulnerable dynamic resource at risk of content theft or enhanced skill squatting (maliciously impersonating a skill [32]), 40% connected to a resource at risk of data leakage, and 10% connected to a resource vulnerable to denial of service or malicious content injection. In addition, we find that dynamic content can evade enforcement mechanisms and avoid state-of-the-art policy enforcement mechanisms such as those discussed in Guo et al. [27]. We give examples of skills that reference vulnerable storage buckets and contain problematic content and demonstrate two ways in which dynamic content can be used to maliciously modify skills or evade policy checks, underscoring the urgent need for stronger security and privacy safeguards within the Alexa skill ecosystem.

1.3 Contributions

This paper presents a comprehensive analysis of Alexa dynamic content, examining its mechanisms, risks, and prevalence. The key contributions of our work are as follows:

- **First Large-Scale Study:** We conducted the first large-scale study on Alexa **dynamic content**, detailing its operational mechanisms and real-world manifestations within the Alexa skill ecosystem. Our analysis covered **10,407** skills, providing new insights into this previously unexamined domain.
- **Novel Risk Identification:** We uncovered several previously unknown risks and attack vectors associated with dynamic content in Alexa skills. Furthermore, we demonstrated successful evasions of content checks using dynamic content resources, highlighting critical vulnerabilities.
- **Skill Interaction Analysis:** Our investigation of skill interactions revealed security and privacy vulnerabilities stemming from dynamic content, leading to unintended skill context switching and the delivery of problematic content.
- **D-Explorer Development:** We designed and implemented D-Explorer, a novel analysis engine comprising:
  - A crawler for skill information and invocation discovery.
  - A ChatGPT-powered chatbot for dynamic content retrieval through skill interactions.
  - Analysis tools for detecting dynamic content vulnerabilities and risks in skill interactions.

2 Background

2.1 The Alexa Skill Ecosystem

The Alexa skill ecosystem allows developers to create custom skills inexpensively and easily, with Amazon providing guidelines, templates, and certification processes [1, 7]. While Amazon has established policies and security requirements, research has shown that policy violating skills can still bypass checks [15, 27], indicating potential vulnerabilities in the certification process. Furthermore, gaps exist in Amazon's guidelines, particularly for skill and S3 bucket naming. Amazon's policy compliance checks occur during skill deployment and re-certification. However, updates to external

resources like Lambda functions or cloud buckets do not trigger re-certification, leading to potential policy violations post-certification.

Alexa skills utilize web resources or cloud buckets for dynamic content. Web resources, accessed through RESTful APIs and Lambda functions, enable interactions with external APIs or websites. Cloud buckets, often used for media files or persistent data, present security risks due to potential over-permissive access controls. Although skill instances have unique identifiers for fine-grained access, these permissions can be difficult to use [5]. Thus, developers often resort to easy-to-implement public read access.

2.2 The Alexa Skill Simulator

The Alexa Developer Portal is used for skill development and testing, including interaction with published skills through a simulator. This simulator provides access to skill conversation trees and limited metadata about attached bucket resources. Although the simulator simplifies automated interactions with Alexa and is a common method of skill interactions in research [27, 34], it does not perfectly replicate real-world device interactions, and some skills may behave differently on physical devices.

3 Dynamic Content Risks and Threat Model

3.1 Content Risks

The inherent unpredictability of Alexa dynamic content opens up users and developers to unanticipated content risks. This is especially the case for skills that reference third-party web resources, such as a news skill pulling from a public RSS feed. As dynamic content changes over time, external events may influence the skill. An example of this is in so-called challenge skills, which retrieve social media challenges for users. However, if the challenge list is not properly curated, harmful challenges may appear. This was the case when an Alexa device told a 10 year old girl to bridge plugged in charger prongs with a penny, a potentially deadly act [48]. Other examples include news skills reporting content that is not appropriate for young children or quote skills that include profanity. The lack of versioning for dynamic content poses a significant challenge, as users cannot easily track changes in resources that may alter a skill's behavior.

Furthermore, dynamic content introduces additional concerns related to Personally Identifiable Information (PII) violations. First, personalized dynamic content incentivizes skill developers to access user information, often in violation of policy. Second, dynamically storing user information allows for saved preferences for convenient usage, often for users who are unaware that third parties are storing this information [40]; this is particularly troubling when considering business skills such as account management or finance tracking. It is important to recognize that malicious content can also be delivered through these same dynamic content mechanisms, not just unanticipated content.

3.2 Resource Risks

Beyond content risks, dynamic content also expands the skill attack surface. Although the skills themselves are not directly vulnerable without compromising developer credentials, external resources do not have the same rigorous security reviews and access controls.

A major risk is misconfigured buckets. Cable et al. [12] found that approximately 10% of cloud storage buckets were misconfigured, with buckets containing public read, write, copy, and delete permissions, and in some cases containing publicly available sensitive information. Buckets with read-only permissions can be exploited to enhance existing skill attacks. Directory listing permissions allow an attacker to discover bucket contents, such as backend code, sensitive information, or audio and image skill resources, which can be used for intellectual property theft or skill squatting. An attacker with delete permissions could launch a denial-of-service (DoS) attack by removing critical resources, such as audio files or configuration data, rendering the skill unusable. Write permissions combined with delete permissions allow an attacker to completely replace skill resources, causing skill disruption, and at worst adding malicious skill behavior, such as asking a user for a credit card number or injecting malicious Alexa commands. Although most skill buckets contain mainly image and audio files, Esposito et al. [24] found that malicious audio files can be used to issue commands to the target device. In addition, audio or image files can still be used to transmit malicious responses or inappropriate content. Thus, vulnerable buckets can lead to privacy breaches or allow for attacks on skills.

Like vulnerable buckets, vulnerable web resources are susceptible to attacks such as denial-of-service (DoS), content modification, and malicious audio streaming. Recent research has shown that up to 95% of websites contain a security vulnerability [19], suggesting a large potential attack surface. Modifications to web resources can directly affect skill behavior. If an attacker gains control of a resource, they can modify skill responses for a variety of purposes, such as impacting quality of service on a skill or modifying responses to provide misinformation for weather or news. Furthermore, the lack of ownership verification for web resources allows for content theft attacks. Attackers can impersonate existing skills to steal revenue, damage reputation, or create new skills using stolen content. An example of this would be a skill that impersonates a blog.

Dynamic content allows malicious developers to bypass Alexa's deployment and immediate post-deployment policy checks, creating an attack vector. By designing compliant initial responses and then injecting malicious content later via RSS feeds, RESTful APIs, or bucket changes, developers can evade restrictions, such as those on skill advertising. Amazon's lack of continuous policy checks prevents the detection and blocking of these exploits.

3.3 Threat Model

We define problematic skill interactions as those where dynamic content leads to the delivery of harmful or inappropriate content, the violation of privacy guidelines, or storage of sensitive data; both intentionally and unintentionally. We consider malicious intent, but do not require it because dynamic interactions cause unintentional harm to users, such as when Alexa encouraged a dangerous internet challenge [48].

For resource-based attacks, we assume that the attacker is not able to change resource permissions or skill code, as such access makes resource based attacks redundant. We assume that the attacker has a basic understanding of how Alexa skills access dynamic

content through Alexa skill tutorials [3–5, 31]. We focus on technical vulnerabilities related to dynamic content, thus excluding social engineering-based exploits.

When considering policy evasions, we assume that the malicious skill provider is attempting to bypass automated and manual checks that Amazon performs. We assume that malicious skill providers own the skills that they are using for evasions and have full access and knowledge of Alexa and AWS resources, such as Lambda functions and S3 buckets.

4 Related Work

Alexa Skill Vulnerabilities and Dangerous Behaviors: Extensive research has explored Alexa skill security and privacy. Le et al. [34] examined skills in the Alexa Kid's category and identified skills exposing children to inappropriate content or collecting personal information, and discovered the risk of "confounding utterances" that could cause unintended skill invocations. Zhang et al. [50] and Kumar et al. [32] examined skill squatting, an attack that leverages invocation uncertainty to create malicious skills that mimic legitimate ones in order to steal data or harm users. Esposito et al. [24] demonstrated how Alexa can be tricked into executing commands from malicious audio played by the device itself, including streamed and downloaded audio. Several researchers have investigated privacy policy compliance and PII handling, using automated tools to detect suspicious behavior and privacy violations [22, 35, 38, 43], finding that Alexa privacy policies are poorly enforced and easy to evade. Talibi et al. [45] developed a user-level defense by enabling runtime detection of malicious PII requests. Ding et al. [21] revealed how malicious skill names can hijack built-in Alexa IOT commands. Comprehensive overviews of Alexa attack surfaces, including skill analysis and external resources, are provided by Leong [37] and Li et al. [39]. However, none of these works consider the impact of dynamic content specifically and none examine third party and remote resources. Furthermore, we expand Le et al's work to examine all skill categories, finding that child related skills exhibit more concerning behaviors than other categories.

Bypassing Skill Verification: Several studies have exposed flaws in skill verification. Cheng et al. [15] and Lentzsch et al. [36] demonstrated the ease of publishing policy-violating skills. Guo et al. [27] developed SkillExplorer, which simulates skill interactions to reveal privacy violations. Edu et al. [23] used machine learning to identify overprivileged skills with broken privacy policies. Hu et al. [29] also explored vulnerabilities in skill verification and methods. Although our work is similar to Guo et al., we specifically consider dynamic content risks and expand our attack surface to consider external resources, which no prior work does.

Vulnerable Cloud Bucket Analysis: Several works have highlighted the security risks of cloud storage buckets. Cable et al. [12] used password generation tools to guess bucket names and identify misconfigured buckets, finding that approximately 10% had vulnerabilities. Continella et al. [16] performed a large scale analysis of vulnerabilities in buckets, finding 191 websites that were vulnerable due to bucket misconfigurations. These works focus on the general bucket ecosystem, and we focus on skill buckets, finding that the skill bucket landscape is significantly more vulnerable.

Alexa Voice Recognition Attacks: Several studies have discovered vulnerabilities in Alexa's voice recognition software, including speech recognition exploits [46, 47] and malicious command execution via hidden or ultrasonic voice commands [13, 49].

5 D-Explorer Design

In this section, we cover the design of D-Explorer and our analysis methodology. Figure 1 shows an overview. D-Explorer is available in its entirety, along with crawler data and sanitized skills results, at https://anonymous.4open.science/r/D-Explorer-7898.

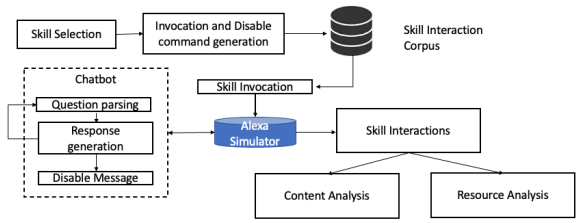


Figure 1: Overall design for D-Explorer.

5.1 Skill Selection and Scope

We examined 10,407 skills from the Alexa ecosystem. We first crawled the Alexa skill store to collect an initial dataset of 59,695 of the most relevant skills in each of the major 22 Alexa skill categories. Although this does exclude least relevant skills in some subcategories, choosing the most relevant skills allows for us to better match consumer choices. Skill information included developer, skill type, ratings, number of reviews, presence of privacy policy/developer policy, description, invocation, and prompts. An example of gathered skill information is found in Appendix B. Although flash briefing skills are dynamic, we removed them from our dataset since flash briefing skills do not contain responses or calls to external resources and are not measurable with our methods. We then removed duplicate skills by checking each skill's unique ID.

From the remaining skills, we choose the top 25% most reviewed skills from each category as an indicator for skill popularity. As we excluded duplicate skills globally, this resulted in an actual percentage of 20% of total skills from our initial set (10,407 skills). Although we considered using skills with low usage numbers, our analysis found that in many cases skills with ratings of 0 or no reviews were non-functional and had never been completed or used. Thus, we examined the skills most relevant to users.

5.2 Command Generation

To automate skill interaction, we began by creating command lists for our chatbot, using skill invocations found in skill details or derived from the skill name. We extracted 0-3 skill commands from the skill page and supplemented them with commands from the skill description using pattern matching. Finally, we added a disable command to mitigate context switching.

Although Cheng et al. [15] indicated that Alexa prevents skill context switching, our interactions revealed instances of context

and intent switching. This occured because skill commands remained accessible across devices after invocation. To mitigate this, we added disable commands to remove skills and limit accidental switching. However, we still encountered switching due to simultaneous usage and failed invocations, potentially impacting our system's performance (discussed in Section 8.2).

5.3 Chatbot Design

We define a request as a command sent to Alexa, a response as the response from Alexa to our command, and an Alexa prompt as a command generated from the skill metadata, instead of interactions.

5.3.1 Developer Portal Interactions. We used the Alexa Developer Portal to interact with Alexa skills through a Selenium webdriver. The chatbot loads the developer portal, sends the skill invocation, gathers a response, and generates another request via the ChatGPT portion of the chatbot. For each request, the chatbot gathers all URLs referenced in the chatbot simulator's HTML to find any potential external resources that were referenced during the interaction. In this manner, the chatbot can gather external resources called for each pair of questions and responses.

5.3.2 Conversation Generation. For each response, D-Explorer first sends a request to ChatGPT's GPT-4.0 model to determine whether the Alexa response is a question or an instruction. If a question is detected, D-Explorer prompts ChatGPT to generate a suitable response (or a list of possible responses). This reduces the risk of unpredictable responses. To create GPT prompts, we iteratively generated a prompt, determined failure cases, and plugged both into the GPT-4.0 model to generate a new prompt. Our finalized prompts generated nearly identical coverage compared to manual skill interactions. These prompts are available at https://anonymous.4open.science/r/D-Explorer-7898/chatbot/gpt_prompts.txt.

To explore multiple conversational branches, D-Explorer tracks a list of possible requests for each response. If that response is revisited, D-Explorer chooses from the list, removing the chosen response as it does. This allows for branch traversal without looping. We implement safeguards in D-Explorer to avoid hang-ups: a maximum conversation depth of 30 for second-level trees, a 300-second skill execution limit, a 120-second prompt duration limit, and a 60-second limit for second-level trees. These limits, set well beyond typical interaction lengths observed in both manual and automated actions, are designed to maximize potential conversation flow while ensuring acceptable performance and mitigating the impact of conversational loops or audio streams, which do not terminate naturally.

5.4 Chatbot Evaluation

5.4.1 Coverage. D-Explorer successfully interacted with 10,407 skills. Of these skills, 6,382 were successfully enabled, 830 were enabled but needed an account linkage for further conversation, and 3,195 were not successfully enabled, primarily due to broken invocations. These are still included in our dataset, as uncertainty from skills not properly enabled can still lead to dynamic content interactions. Excluding Amazon runtime generated resources, there were 6,971 unique links to external content, with 2,018 unique top-level paths and 863 unique domains.

To measure skill interaction coverage coverage, we analyzed collected skill interactions in the same way as Le et al. [34], with 4 criteria including the number of unique responses from Alexa, the maximum depth in a prompt tree, the maximum branches in a prompt tree and the number of initial utterances. Collectively, Figure 2 shows our performance in each of these categories. In each category, we outperform Le et al., showing we elicit more unique responses, conversational branches, and converse longer than prior works. We also generate more initial utterances, which allows us to more accurately match common user behavior.

5.4.2 Performance. Although we did not conduct a detailed performance analysis, we were able to run each skill interaction with an estimated average time of 101.32 seconds per skill, running ten chatbot instances in parallel on a Macbook Pro with a 2.6 GHz 6 Core Intel I7 processor. In comparison, Skillbot (Le et al, [34]) took an estimated 183.714 seconds per skill, and SkillExplorer (Guo et al, [27]) took 627 seconds on average. Furthermore, only D-Explorer waits until all external resources are loaded for each request, which significantly increased overhead. The increased performance and coverage of our system showcases the potential of LLMs for conversational analysis projects.

5.5 Analysis

We conducted automated analysis on the chatbot results to assess the impact of dynamic content risks on deployed Alexa skills. For our keyword matching classifiers, we tested on a ground truth dataset of 1000 interaction lines randomly sampled from the interaction dataset. We consider the following categories:

5.5.1 Profane content. We define profane content as any content that is offensive or inappropriate. This includes hate speech, profanity, and sexual content. We use keyword matching or existing classifiers to find profane content, and determine if any matching content is dynamic through the dynamic content detection tool below.

We used the hateBERT classifier [14] to search for hate speech. The hateBERT classifier exhibits F1 scores of anywhere from 0.75 to 0.52 on unfamiliar data. Although these scores are low, we did not detect any hate speech through keyword matching and we believe that there are no examples in our interaction dataset.

We used the profanity_check [51] Python library to detect profanity in the interaction dataset. This library uses machine learning to detect profanity with a claimed F1 score of 0.88.

The initial F1 score against the ground truth was .5, a result primarily influenced by the low number of True Positives (n=2). Both single true positive and false negative were context dependent, and beyond our classifier's ability. Due to the low number of matches for possible inappropriate content, we manually verified all responses found in our keyword checks, although the possibility of false negatives remains.

5.5.2 Requests for Personal Information. We define requests for personal information as any skill that asks for Personally Identifiable Information (PII), especially against policy guidelines or when the skill does not reference a need for such information in its description. We use keyword matching and/or existing classifiers to

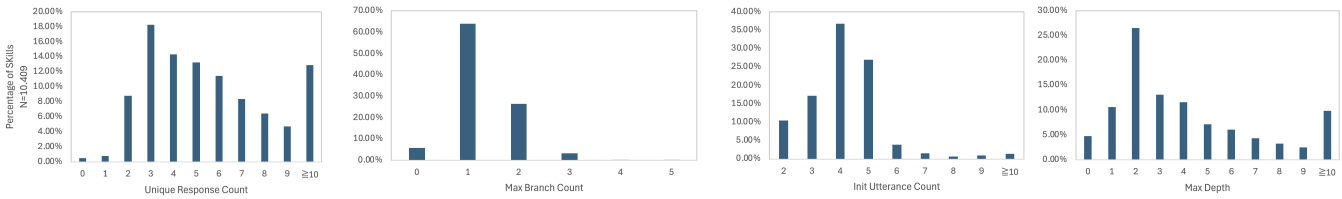


Figure 2: Chatbot performance metrics.

find inappropriate content and determine if any matching content is dynamic.

We generated a corpus of privacy-related keywords based on the U.S. Department of Defense's controlled unclassified information PII keywords list [10]. This list is then augmented with keywords that indicate violations of Alexa policy guidelines.

Our evaluation against the ground truth dataset yielded an F1 score of 0.4, primarily attributable to a high number of false positives associated with zip codes. While this initial result was concerning, further analysis revealed a significantly higher prevalence of zip codes found within the ground truth dataset compared to the full dataset. To address this, we conducted a second ground truth evaluation on an additional 500 randomly chosen interactions, resulting in an F1 score of 0.8, and measured 50 PII-positive interactions from the full dataset, finding a precision of 80%.

We also used Presidio, a python-based PII recognizer developed by Microsoft [41], to detect PII. To reduce the risk of false positives, we ignored any detected instances that have a probability of less than 100%. However, despite this, the Presidio tool is much broader when classifying PII than our keyword-based tool, and the results of the Presidio analysis are reported separately.

5.5.3 Dynamic Content. We generated a corpus of words that define dynamic content through interactions with skills, since no existing corpus exists for this content. We included temporal keywords, as well as keywords involving recent events and keywords describing items and functions that require a temporal element to work properly. We refined this corpus by manually analyzing the skills and eliminating any word or phrase that results in a high number of false positives (greater than 80%). This specific threshold was chosen as to remove false positives without significantly increasing false negatives. Our ground truth evaluation resulted in an F1 score of .75.

5.5.4 Bucket Usage. We searched responses for third-party resources. We excluded text-to-speech resources and CAPS-SSE buckets, as those are generated at runtime using Amazon services.

We gathered unique bucket URLs and determine permission status of the identified buckets by trying to interact with the contents of the bucket using a standard web browser and through the Amazon boto3 library. Although there are a few buckets within our dataset that are not CloudFront or S3 buckets, these are not reported in the bucket usage results due to the difficulty of distinguishing them from regular websites without manual analysis. Although similar works [12] do not consider public read, public read access conflicts with existing security guidelines and resource usage tutorials for Alexa skills [5, 9], and are thus misconfigured. Configuration types are classified as follows:

- **List and Read:** Buckets that allow all contents of the bucket to be listed and read by the public.
- **Read:** Buckets that expose specific resource files but do not allow bucket content lists and tags to be read.
- **Write:** Buckets that allow new files to be written or existing files to be overwritten.
- **Delete:** Buckets that allow files to be deleted.
- **Secured:** Buckets that prevent file access and bucket read.

5.5.5 Web Based Exploits. We demonstrate potential web-based exploits by conducting tests with our own skill and S3 bucket. We do not examine all retrieved sites for vulnerabilities ourselves, but note that existing studies and CVEs cover web security or exploitable content. Of note are skills pulling audio data from radio websites, which is a potential exploitation vector [24].

6 Evaluation and Results

In our analysis, we answer the following research questions:

- RQ 1: What are the content risks in the Amazon Alexa ecosystem?
- RQ 2: What are the resource risks in the Alexa skill ecosystem?

6.1 What are the content risks in the Amazon Alexa ecosystem?

6.1.1 PII. Of the skills examined, 3,572 (34%) contained dynamic content in conversations. We found 797 skills (22% of skills with dynamic content, 8% of the total) with problematic dynamic content (content that is potentially inappropriate for children, offensive or asks for PII). Of these, 483 requested PII. Many PII interactions occurred when PII from an Alexa user was used to generate dynamic content without permission (a policy violation). In many cases, a skill responded with an interaction detailing location or distance from the Alexa user, despite no requests for that data. In contrast, several skills asked for permission and were not able to continue, demonstrating correct behavior. One skill used stored credit card information to order a book from the Amazon store. Alexa certification procedures should detect these issues, though they do not [15]. As account linkage is implemented in skill code, changes to backend resources cannot change a skill to ask for Alexa account information post-certification. In general, we see that dynamic content drives developers to implement skills that can violate privacy policies.

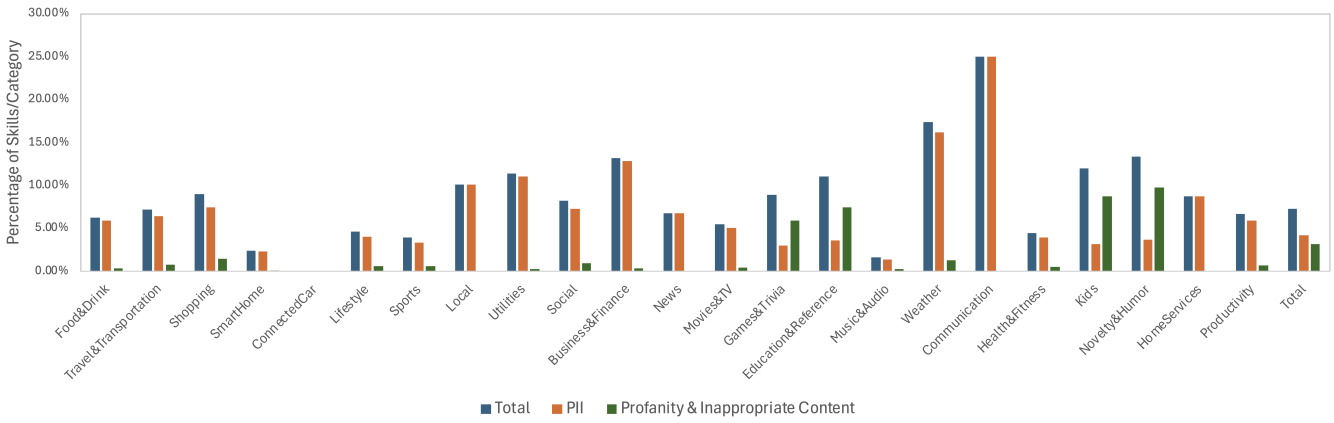


Figure 3: Percentage of skills with problematic content.

Many skills indicated through conversation that they stored PII. Most of these did not link to the Alexa account and therefore stored persistent data. This implies that any data stored are stored in the skill backend, either on a website or in a skill bucket. This exposes users to potential information theft, as many backend resources are poorly secured and allow all contents to be read. Furthermore, two skills wrote Alexa information to these resources. Examples of stored PII include addresses, emails, phone numbers, and birthdays. One skill stored and tracked credit card information, including credit card numbers, and several banking skills interacted with bank account information. The credit card tracker skill referenced an unsecure bucket, although that bucket fortunately did not contain stored credit card information. Unfortunately, many Alexa users do not know or understand that these skills are third-party and that the data collected is not protected by Amazon [40]. Although Amazon invests heavily in security [2, 42], third party skill developers, like those of the credit card tracker, are likely not able to do the same [25]. Concerningly, 12% of Business & Finance skills exhibited these behaviors - a category where unsecurely stored data or malicious data requests can significantly harm a user.

As shown in Figure 3, many PII requests are found in categories that use PII data, such as location data for weather, restaurants or shopping skills, or financial information in business skills. However, several Education & Kids skills ask for birthdays or names. In some cases (for example, a historical date skill) this data was not needed. Although an adult may think twice about answering a request for a birthday or address, a child may not understand that a skill that makes animal sounds does not need an address.

Including Presidio results, 5,120 skills (49%) contain problematic dynamic content, with 4,766 of those flags originating from Presidio. This highlights the need for further research to determine the accuracy of this tool and other existing tools specifically for Alexa content.

6.1.2 Profanity and Inappropriate Content. We identified 292 files containing profanity or profane language due to dynamic content, and 38 files with inappropriate content. In total, 83% of the skills that contained profanity and 76% of the files that contained inappropriate content were in categories commonly used by children, the exact type of users to whom this content is most dangerous. Figure 3 shows the distribution of skills per category, with the highest percentage of skills that serve inappropriate content found in the categories Kids, Novelty & Humor, Education & Reference, and Games & Trivia. In these categories, 8% of the skills contained profanity or inappropriate content, compared to 3% of all categories. Many of the inappropriate content interactions resulted from an unexpected interaction with the "Alexa, what is the score" command found in flashcard and quiz skills in these categories. This command returned a news article on a Welshman who drew the largest GPS drawing of a penis while on foot. While not inherently dangerous, it is concerning that this seemingly benign interaction returns an article to which parents might not want their children to be exposed. Another skill, this time in a math tutor skill, context-switched to a different skill during normal interaction and returned a news article about a man arrested for a striptease at a restaurant. In the Lifestyle category, a Bible study podcast skill proceeded to play podcasts on cannibals, concubines, and sex cults when there were no Bible study podcasts to stream. Other examples include several travel skills that told inappropriate jokes, such as "Yo mama's so fat she needs a GPS to find her a** h***" or "Yo mama's so poor she chases a garbage truck with a grocery list". Fortunately, although we found several challenge skills, none of them referenced dangerous challenges that we know of, although the possibility of dangerous challenges remains. These interactions highlight dynamic content risks, most worryingly in skills used by children.

6.2 What are the resource risks in the Alexa skill ecosystem?

To examine the resource risks within the Alexa ecosystem, we focused on calls to external resources during interactions. We excluded links consisting of CAPS-SSE and tinytts/tinyaudio references, as these are generated by Amazon during runtime and are not exploitable, as they expire shortly after use. We also excluded 11 domains that are present in more than 90% of skills and any generic S3 domains, as these domains are, to our knowledge, related to the developer portal or to generic Amazon operations and thus outside the scope of our study. Excluding these resources, there are

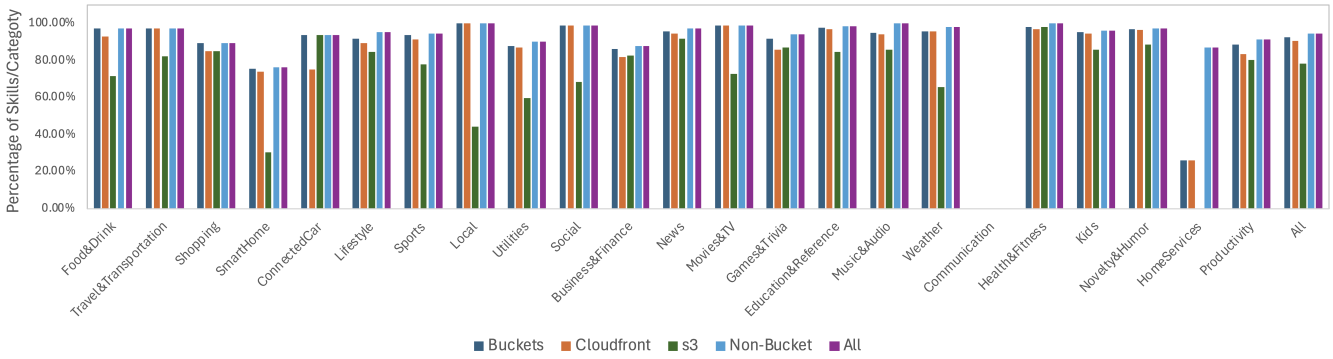


Figure 4: Percentage of skills that use external resources per category.

4,890 unique links to external content, with 1,805 unique top-level paths and 810 unique domains. URLs were categorized into links, domains, and top-level folders to distinguish domains that expose specific resources, domains that expose multiple resources through resource leakage, and domains that are not file hosting sites, but rather login pages or skill websites.

Most referenced files are image and audio files used during Alexa interactions, such as a recording of whale sounds or an image of the latest special at a restaurant. Although most skills take advantage of various Amazon text-to-speech offerings to provide audio content, some skills depend solely on stored S3 bucket audio. These skills avoid the costs associated with tinytts and tinyaudio, but are more vulnerable to content modification, deletion, and theft through insecure S3 buckets. In general, the requested resources are split primarily between image and audio files, with 49% image, 37% audio, and 14% other. See Appendix A for more details on file usage. Interestingly, 95% of the skills referenced a web resource of some kind. These resources are broken down into total buckets, S3 buckets, CloudFront buckets, and web resources by skill category in Figure 4.

6.2.1 Bucket Risks. We found that 9,629 skills utilize at least one resource bucket, with 8,143 skills referencing an S3 bucket and 9,421 referencing a CloudFront bucket. These skills used 365 unique buckets, with a few buckets providing resources for numerous skills, such as a bucket referenced in interactions with 1,970 skills. Although no skills shared a developer, we found that many skills share buckets, resulting in a high number of skills using a single bucket. Of the buckets found, 23% were hosted by CloudFront, and 77% by Amazon S3, with 84% misconfigured (Table 1). Interestingly, as shown in Figure 4, although most of the identified buckets are S3 buckets, more skills use CloudFront-hosted buckets. Of the skills examined, 90% skills reference buckets that are partially exposed to public access (Table 2), 40% reference buckets with full read access to their contents, and 10% reference buckets that allow public upload and deletion of files. Most fully exposed buckets are S3 buckets, even though more skills reference CloudFront buckets. This is not surprising, as CloudFront buckets can have enhanced security features that S3 buckets do not. Furthermore, the percentage of poorly configured buckets, more than 84%, is eight times higher than observed in previous S3 bucket configuration studies [12]. This discrepancy is attributed to the added complication of Alexa bucket interactions, which must allow access from Alexa devices and skills but prevent bucket access from the general public. We also observed that major providers tend to have fully secured buckets, although exceptions exist, such as a bucket containing 1,000 copyrighted audio files critical to the skill's function. Furthermore, our results include many audio streaming files that have been noted as vectors for Alexa command injection. As discussed in Section 3.2, these unsecure buckets present risks ranging from content theft to malicious command injection to skill DoS, in the case of skills that reference buckets with read/write access.

Read Permissions Only: Buckets with file read permissions only constitute the vast majority of buckets, both in terms of individual buckets and skill references. These buckets expose specific skill files, but do not allow for information leakage through list interactions or bucket modification. These files are primarily exploitable through file theft, either to simply steal copyrighted information or resources, or to execute skill squatting attacks. Several CloudFront buckets, for example, reference multiple skill image and JavaScript files. One bucket appears to serve all of the application code and CSS files to a financial skill during skill interaction. Audio files can be exploited in the same manner; one skill, for example, serves entire 45-minute podcast episodes that can be stolen or used in skill squatting. Other examples include a sleep sounds skill that serves images and audio files and a financial skill that serves skill icons and audio.

Table 1: Bucket Vulnerabilities by Provider

Bucket Provider	Secured	Read	List & Read	Write & Delete	Total
S3	55	197	27	3	282
CloudFront	5	65	12	N/A	83
Total	60	262	39	3	365

Buckets With List & Read Permissions: Buckets with list permissions expand the attack surface by introducing information leakage to content theft. First, attackers can expand skill squatting attacks or content theft to multiple skills, if a referenced bucket

Table 2: Bucket Vulnerabilities by Skill

Bucket Provider	Secured	Read	List & Read	Write & Delete	Total
S3	2512	7897	2265	998	10407
CloudFront	1413	8996	5786	0	10407
Total	1045	9362	3229	998	10407

serves multiple skills. Second, an attacker might find private user information such as stored card numbers from a credit card tracker or a birth date stored in an alcohol selection skill. As an example of both of these risks, a bucket with read permissions was referenced in 1,970 skill conversations and contained over 100 audio and image files and over a thousand total files. These included API keys, skill IDs, and compressed data in text files. The bucket was active at the time of writing. Another active bucket contained hundreds of audio and image files, exceeding a thousand files in total. A bucket referenced by several skill privacy policies contained .apk files, 252 instances of JavaScript code files, 297 CSS files, 43 HTML files, and log files from skill instances. Other buckets contained Apple and Android application files. Additional information found in Alexa skill buckets included references to customers and advertisers, company documentation, backups, or even S3 buckets of the owner's customers (fortunately, these were secured).

Buckets with Write & Delete: The most exploitable buckets are those with write & delete permissions. We were only able to scan S3 buckets due to the obfuscated nature of CloudFront buckets, which rendered them incompatible with our tools. We identified three buckets with public read and write permissions, all of which contained primarily image and/or audio files, although one also contained company documentation. These skills are most vulnerable to attacks, as an attacker can download a file, modify it, and upload it to affect skill performance or inject malicious content. Without access to a skill's code, it is difficult for an attacker to change skill behavior enough to steal information, at least in the cases we observed. However, ruining the reputation of the skill, causing harm for fun, or even simply deleting the contents of the bucket to initiate a DoS attack on the skill are all much more trivial. As these buckets served 10% of all skills, the potential impact is substantial.

Content Evasion: Finally, we note that all skills that utilize buckets can evade content checks. Although skill buckets do not host deployed code, a malicious skill developer can easily modify and re-upload a file to cause different behavior. For example, an audio command that initially requests and stores a favorite song could be changed to request and store an address, name, or birthday. Some of the skills mentioned in Section 6.1 may request PII due to this evasion.

Proof of Concept Attack: We conducted a proof-of-concept of attacks on write/delete buckets as follows. First, we create a bucket with full read, write, delete, and list permissions. We associated a skill with an item in this bucket and using public access, deleted the audio resource, and replaced it with a modified version. This modified file could request PII or contain problematic content. In our case, it requested PII. We also demonstrated that simply deleting the resource negatively affects skill performance. Note that both attacks can be used as content evasion mechanisms. This attack also avoids new verification checks [36].

6.2.2 Web Resource Usage. Overall, 9,855 skills reference at least one nonbucket resource. These included 3,017 unique URLs, spanning 445 unique domains. Many of these domains were websites hosting a resource for the skill, such as image, audio, or even CSS files, but in most cases these resources were hosted behind a traditional site login page, rather than S3-style buckets. Unlike the bucket domains, many resources pulled from these web domains were .aac or .ashx files, which were exclusively used to stream audio content (typically radio stations or podcasts).

Attack Vectors: While not as immediately exploitable as bucket resources, web resources present several attack vectors. Content theft remains a concern, with 83% of domains allowing low-level file access, comparable to buckets with read access. More skills refer to web endpoints than bucket endpoints (Figure 5), predominantly for audio/image files and some code (Appendix A).

Although direct data leakage is low risk—only 31% of URLs allow top-level access (often false positives), and 56% of accessible domains point to home/login pages—web resources suffer from various other vulnerabilities. Examples impacting skills include a 2023 IceCast/WordPress vulnerability allowing radio player deletion (potential DoS) [18], a 2017 image hosting site flaw enabling arbitrary file uploads [17], and a 2024 Digital Ocean OpenSSH vulnerability allowing Linux system takeovers (potentially leading to complete skill modification) [20].

Evasions & Proof of Concept Attack: Web resources, as with bucket resources, can be used to bypass policy checks and inject malicious or restricted content. To demonstrate this, we performed a proof-of-concept web modification evasion using RSS feeds. We first set up a blog website that published an RSS feed of articles. We then created a skill that reads the feed and informs the user of the top 3 articles on the page. Once the skill is running, we replaced the top 3 articles with malicious content. This attack can be used to insert advertisements into skill behavior, replace responses to elicit personal information or cause skill degradation, or DoS. With this attack, the skill does not need to pass a new verification check [36].

7 Defenses and Mitigations

User-side content moderation tools, like Talebi et al. [45], effectively prevent PII leakage and could easily expand to cover other problematic responses, including image and streamed audio, mitigating many dynamic content and interaction risks.

To address context switching and unpredictable interactions, Amazon must remove broken or inactive skills, removing much of the ambiguity that causes skill switching or unexpected dynamic content. A dedicated study on skill deprecation is needed to determine the percentage of nonfunctional skills. As of the writing of this paper, it appears that skills can potentially remain in the Alexa ecosystem indefinitely.

Finally, dynamic content warnings are completely inadequate. While present in skill descriptions, they lack crucial details (content source, changes, usage) and are absent during audio activation. Improving dynamic content warnings and adding them to audio invocations is critical.

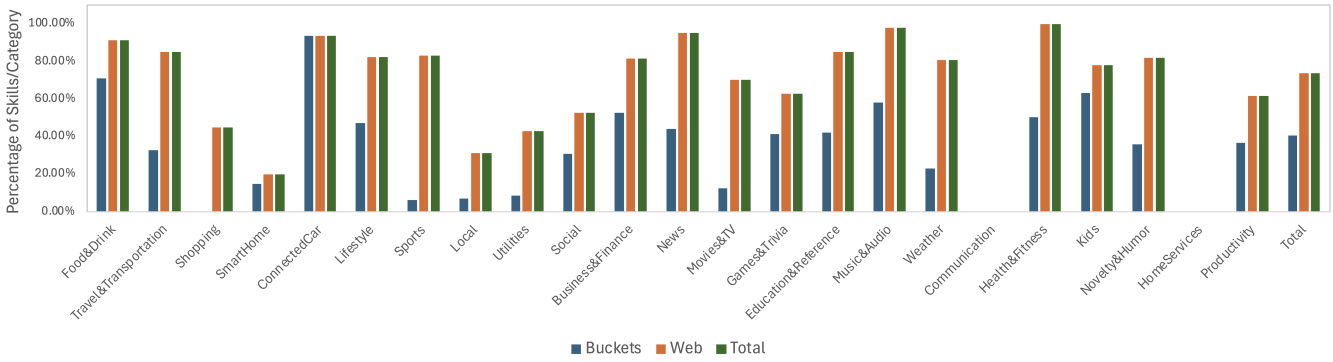


Figure 5: Percentage of skills that reference vulnerable file endpoints.

8 Discussion

8.1 Ethical Considerations

Scanning buckets raises ethical concerns. To test permissions (read, write, delete, list), we created, wrote, and deleted temporary files. Unlike Cable et al. [12], we also tested delete permissions and briefly examined downloaded files before deletion. We used Amazon's boto3 library with HTTPS encryption and will not publish any list of vulnerable buckets or related information.

Additional considerations include the actual chatbot interactions themselves. As we used a real Alexa account when interacting with skills, many skill interactions contain identifiable or protected data. As a result, we do not publish the skill interactions.

We informed the owners of the vulnerable buckets found to the best of our ability and have reached out to Amazon. We were able to contact 41 bucket owners, corresponding to 46 buckets total. We have received acknowledgments of the issue from 4 bucket owners at the time of writing.

8.2 Limitations and Future Work

Our methodology could not capture references in obfuscated device code, and we found that traffic analysis via the simulator was ineffective for this issue, requiring real world interactions. However, we believe our approach is more comprehensive and scalable.

As noted in Section 5.2, chatbot interactions with the Alexa simulator showed context switching. While our skill list is recent, it cannot be fully current due to the dynamic nature of the Alexa skill ecosystem. Consequently, chatbot interactions can be unpredictable due to context switching or outdated invocations/disable commands. Further study on Alexa invocation and interaction unpredictability is needed.

We analyzed a subset of the most relevant skills in the Alexa skill ecosystem. Thus, there is the potential that our subset does not reflect the general ecosystem. Furthermore, although our chatbot is designed to elicit hidden skill responses, the chatbot is not guaranteed to elicit 100% of all responses for all skills.

We observed that many skills referenced the same buckets, even when those skills claimed different developers. Although some of this behavior may be due to context switching, we verified that many skills share buckets. A further study is needed to determine the exact extent of bucket sharing between skills, and the related impacts on bucket and skill security.

Although we performed in-depth analysis of bucket resources, we did not perform the same level of analysis for web resources, due to the additional complexity of this task. Future work could include a large-scale vulnerability analysis of web resources used by Alexa skills to address this disparity.

Finally, our identification of problematic content and dynamic content is basic. We intend to improve these detection classifiers by experimenting with machine learning resources, including ChatGPT, to perform these detection tasks.

9 Conclusion

Our examination of dynamic content has revealed alarming security issues with the skill ecosystem, finding that 7% of skill interactions contain problematic content and that a staggering 90% of skills reference an unsecured resource. Inappropriate dynamic content is a significant issue, with 8% of skills containing inappropriate dynamic interactions and 92% of the total inappropriate interactions found in categories commonly used by children. Furthermore, 12% of Business and Finance skills mishandled PII (Personally Identifiable Information), storing it in third party web or bucket resources, a practice which many users are not aware of.

Other skills (over 10% in categories such as Weather, Utilities, and Local) used location data without permission or asked for and stored location data. Skills using bucket resources, in particular, are particularly vulnerable to data theft and modification, with 40% of skills referencing buckets open to data theft and 10% referencing buckets vulnerable to resource changes and deletions. While web content is less likely to leak data, web content provides an avenue for censor evasion as well as exposing a skill to the web attack surface. Dynamic content labels are woefully insufficient and do not reflect the risks of dynamic content, even when they are seen.

To solve these issues, developers and users must advocate for increased content management regulations and user protections. Furthermore, it is critical that developers implement the correct security policies in dynamic resources to prevent data theft and malicious interactions. Only through rigorous regulation and vigilant developer practices can we ensure that the Alexa skill ecosystem fulfills its potential without compromising user security and safety.

References

[1] Amazon. 2022. Alexa Skill Blueprints. https://blueprints.amazon.com

[2] Amazon. 2023. *2023 Amazon Annual Report.* https://s2.q4cdn.com/299287126/files/doc_financials/2024/ar/Amazon-com-Inc-2023-Annual-Report.pdf

[3] Amazon. 2024. *Beginner's guide to building Alexa skills.* https://developer.amazon.com/en-US/alexa/trainings_and_workshops

[4] Amazon. 2024. *Use Media Files with Your Alexa-Hosted Skill.* https://developer.amazon.com/en-US/docs/alexa/hosted-skills/alexa-hosted-skills-media-files.html

[5] Amazon. 2024. *Use Personal AWS Resources with Your Alexa-Hosted Skill.* https://developer.amazon.com/en-US/docs/alexa/hosted-skills/alexa-hosted-skills-personal-aws.html

[6] Amazon. 2025. Alexa Skills Kit. https://developer.amazon.com/en-US/alexa/alexa-skills-kit

[7] Amazon. 2025. Certify and Publish Your Skill. https://developer.amazon.com/en-US/docs/alexa/certify/certify-your-skill.html

[8] Amazon. 2025. Create Alexa Skills Kit | Amazon Alexa Voice Development. https://developer.amazon.com/en-US/alexa/alexa-skills-kit

[9] Amazon. 2025. *Security best practices for Amazon S3.* https://docs.aws.amazon.com/AmazonS3/latest/userguide/security-best-practices.html

[10] National Archives. 2021. CUI Category: Sensitive Personally Identifiable Information. https://www.archives.gov/cui/registry/category-detail/sensitive-personally-identifiable-info

[11] Umair Bashir. 2025. *Smart speaker ownership by brand in the U.S. as of December 2024.* https://www.statista.com/forecasts/997149/smart-speaker-ownership-by-brand-in-the-us#:~:text=Published%20by,consumers%20in%20the%20United%20States.

[12] Jack Cable, Drew Gregory, Liz Izhikevich, and Zakir Durumeric. 2021. Stratosphere: Finding Vulnerable Cloud Storage Buckets. In *24th International Symposium on Research in Attacks, Intrusions and Defenses (RAID '21)* (San Sebastian, Spain) *(RAID '21)*. ACM, New York, NY, USA, 399–411. doi:10.1145/3471621.3473500

[13] Nicholas Carlini, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, and Wenchao Zhou. 2016. Hidden Voice Commands. In *Proceedings of the 25th USENIX Conference on Security Symposium* (Austin, TX, USA) *(SEC'16)*. USENIX Association, USA, 513–530.

[14] Tommaso Caselli, Valerio Basile, Jelena Mitrović, and Michael Granitzer. 2021. HateBERT: Retraining BERT for Abusive Language Detection in English. arXiv:2010.12472 [cs.CL] https://arxiv.org/abs/2010.12472

[15] Long Cheng, Christin Wilson, Song Liao, Jeffrey Young, Daniel Dong, and Hongxin Hu. 2020. 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* (Virtual Event, USA) *(CCS '20)*. Association for Computing Machinery, New York, NY, USA, 1699–1716. doi:10.1145/3372297.3423339

[16] Andrea Continella, Mario Polino, Marcello Pogliani, and Stefano Zanero. 2018. There's a Hole in That Bucket! A Large-Scale Analysis of Misconfigured S3 Buckets. In *Proceedings of the 34th Annual Computer Security Applications Conference* (San Juan, PR, USA) *(ACSAC '18)*. Association for Computing Machinery, New York, NY, USA, 702–711. doi:10.1145/3274694.3274736

[17] CVE Details. 2017. CVE-2017-15962. https://www.cvedetails.com/cve/CVE-2017-15962/ Accessed: 2025.

[18] cve.org. 2023. CVE-2023-4024. https://www.cve.org/CVERecord?id=CVE-2023-4024 Accessed: 2025.

[19] Nurullah Demir, Tobias Urban, Kevin Wittek, and Norbert Pohlmann. 2021. Our (in)Secure Web: Understanding Update Behavior of Websites and Its Impact on Security. In *Passive and Active Measurement: 22nd International Conference, PAM 2021, Virtual Event, March 29 – April 1, 2021, Proceedings* (Cottbus, Germany). Springer-Verlag, Berlin, Heidelberg, 76–92. doi:10.1007/978-3-030-72582-2_5

[20] DigitalOcean. 2024. Regresshion vulnerability: Recommended actions and steps we've taken. https://www.digitalocean.com/blog/regresshion-vulnerability-recommended-action Accessed: 2025.

[21] Wenbo Ding, Song Liao, Long Cheng, Xianghang Mi, Ziming Zhao, and Hongxin Hu. 2024. Command Hijacking on Voice-Controlled IoT in Amazon Alexa Platform. In *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security* (Singapore, Singapore) *(ASIA CCS '24)*. Association for Computing Machinery, New York, NY, USA, 654–666. doi:10.1145/3634737.3657010

[22] Jide Edu, Xavier Ferrer-Aran, Jose Such, and Guillermo Suarez-Tangil. 2022. Measuring Alexa Skill Privacy Practices across Three Years. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 670–680. doi:10.1145/3485447.3512289

[23] Jide S. Edu, Xavier Ferrer-Aran, Jose Such, and Guillermo Suarez-Tangil. 2023. SkillVet: Automated Traceability Analysis of Amazon Alexa Skills. *IEEE Transactions on Dependable and Secure Computing* 20, 1 (Jan. 2023), 161–175. doi:10.1109/tdsc.2021.3129116

[24] Sergio Esposito, Daniele Sgandurra, and Giampaolo Bella. 2022. Alexa versus Alexa: Controlling Smart Speakers by Self-Issuing Voice Commands. arXiv:2202.08619 [cs.CR] https://arxiv.org/abs/2202.08619

[25] Christophe Foulon. 2025. *Small Business Cyber Attack Statistics: A Wake-Up Call for SMBs.* https://www.cisoplatform.com/profiles/blogs/small-business-cyber-attack-statistics-a-wake-up-call-for-smbs-ch#:~:text=%2D%20Only%2014%25%20of%20SMBs%20have,an%20attack%20to%20purchase%20coverage

[26] Google. 2025. Prepare your app for review - Play Console Help. https://support.google.com/googleplay/android-developer/answer/9859455?hl=en

[27] Zhixiu Guo, Zijin Lin, Pan Li, and Kai Chen. 2020. SkillExplorer: Understanding the Behavior of Skills in Large Scale. In *Proceedings of the 29th USENIX Conference on Security Symposium*. USENIX Association, USA, 18 pages. https://www.usenix.org/conference/usenixsecurity20/presentation/guo

[28] D. Harel. 2022. *The Smart Audio Report.* Technical Report. National Public Media, Chicago, IL, USA.

[29] Hang Hu, Limin Yang, Shihan Lin, and Gang Wang. 2020. Security Vetting Process of Smart-home Assistant Applications: A First Look and Case Studies. arXiv:2001.04520 [cs.CR] https://arxiv.org/abs/2001.04520

[30] Apple Inc. 2025. App Review - App Store - Apple Developer. https://developer.apple.com/app-store/review/

[31] Justin Jeffress. 2017. *Add Dynamic Content to Your Skill to Keep Users Engaged over Time.* Amazon. Retrieved Sep 1, 2022 from https://developer.amazon.com/blogs/alexa/post/8be10dbe-48e4-4d7c-9638-8657ffdbeb9e/add-dynamic-content-to-your-skill-to-keep-users-engaged-over-time

[32] Deepak Kumar, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, and Michael Bailey. 2018. Skill squatting attacks on amazon alexa. In *Proceedings of the 27th USENIX Conference on Security Symposium* (Baltimore, MD, USA) *(SEC'18)*. USENIX Association, USA, 33–47.

[33] Federica Laricchia. 2022. *Total number of Amazon Alexa skills from January 2016 to September 2019.* https://www.statista.com/statistics/912856/amazon-alexa-skills-growth/

[34] Tu Le, Danny Yuxing Huang, Noah Apthorpe, and Yuan Tian. 2022. SkillBot: Identifying Risky Content for Children in Alexa Skills. *ACM Trans. Internet Technol.* 22, 3, Article 79 (jul 2022), 31 pages. doi:10.1145/3539609

[35] Tu Le, Dongfang Zhao, Zihao Wang, XiaoFeng Wang, and Yuan Tian. 2024. Alexa, is the skill always safe? Uncover Lenient Skill Vetting Process and Protect User Privacy at Run Time. In *Proceedings of the 46th International Conference on Software Engineering: Software Engineering in Society* (Lisbon, Portugal) *(ICSE-SEIS'24)*. Association for Computing Machinery, New York, NY, USA, 34–45. doi:10.1145/3639475.3640102

[36] Christopher Lentzsch, Sheel Jayesh Shah, Benjamin Andow, Martin Degeling, Anupam Das, and William Enck. 2021. Hey Alexa, is this Skill Safe?: Taking a Closer Look at the Alexa Skill Ecosystem. In *Proceedings of the 28th ISOC Annual Network and Distributed Systems Symposium (NDSS)*.

[37] Raphael Leong. 2017. *Analyzing the Privacy Attack Landscape for Amazon Alexa Devices.* Master's thesis. Imperial College London.

[38] Song Liao, Long Cheng, Haipeng Cai, Linke Guo, and Hongxin Hu. 2023. SkillScanner: Detecting Policy-Violating Voice Applications Through Static Analysis at the Development Phase. In *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security* (Copenhagen, Denmark) *(CCS '23)*. Association for Computing Machinery, New York, NY, USA, 2321–2335. doi:10.1145/3576915.3616650

[39] Yanyan Lit, Sara Kim, and Eric Sy. 2021. A Survey on Amazon Alexa Attack Surfaces. In *2021 IEEE 18th Annual Consumer Communications & Networking Conference (CCNC)* (Las Vegas, NV, USA). IEEE Press, 1–7. doi:10.1109/CCNC49032.2021.9369553

[40] David Major, Danny Yuxing Huang, Marshini Chetty, and Nick Feamster. 2021. Alexa, Who Am I Speaking To?: Understanding Users' Ability to Identify Third-Party Apps on Amazon Alexa. *ACM Trans. Internet Technol.* 22, 1, Article 11 (Sept. 2021), 22 pages. doi:10.1145/3446389

[41] Microsoft. 2025. Presidio: Data Protection and De-identification SDK. https://microsoft.github.io/presidio/

[42] Mark Ryland. 2023. *Our commitment to shared cybersecurity goals.* https://aws.amazon.com/blogs/security/our-commitment-to-shared-cybersecurity-goals/

[43] Hassan A. Shafei, Hongchang Gao, and Chiu C. Tan. 2024. Measuring privacy policy compliance in the Alexa ecosystem: In-depth analysis. *Comput. Secur.* 144, C (Sept. 2024), 15 pages. doi:10.1016/j.cose.2024.103963

[44] Satyajit Sinha. 2024. *State of IoT 2024: Number of connected IoT devices growing 13% to 18.8 billion globally.* https://iot-analytics.com/number-connected-iot-devices/

[45] Seyed Mohammadjavad Seyed Talebi, Ardalan Amiri Sani, Stefan Saroiu, and Alec Wolman. 2021. MegaMind: a platform for security & privacy extensions for voice assistants. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services* (Virtual Event, Wisconsin) *(MobiSys '21)*. Association for Computing Machinery, New York, NY, USA, 109–121. doi:10.1145/3458864.3467962

[46] Tavish Vaidya, Yuankai Zhang, Micah Sherr, and Clay Shields. 2015. Cocaine Noodles: Exploiting the Gap between Human and Machine Speech Recognition. In *Proceedings of the 9th USENIX Conference on Offensive Technologies* (Washington,

D.C.) *(WOOT'15)*. USENIX Association, USA, 16.
[47] Payton Walker, Nathan McClaran, Zihao Zheng, Nitesh Saxena, and Guofei Gu. 2022. BiasHacker: Voice Command Disruption by Exploiting Speaker Biases in Automatic Speech Recognition. In *Proceedings of the 15th ACM Conference on Security and Privacy in Wireless and Mobile Networks* (San Antonio, TX, USA) *(WiSec '22)*. Association for Computing Machinery, New York, NY, USA, 119–124. doi:10.1145/3507657.3528558
[48] Derek Wise. 2021. *Amazon responds after Alexa encourages dangerous penny challenge.* https://9to5mac.com/2021/12/28/amazon-responds-after-alexa-encourages-dangerous-penny-challenge/
[49] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. DolphinAttack: Inaudible Voice Commands. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (Dallas, Texas, USA) *(CCS '17)*. Association for Computing Machinery, New York, NY, USA, 103–117. doi:10.1145/3133956.3134052
[50] Nan Zhang, Xianghang Mi, Xuan Feng, XiaoFeng Wang, Yuan Tian, and Feng Qian. 2019. 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)*. 1381–1396. doi:10.1109/SP.2019.00016
[51] Victor Zhou. 2019. Building a Better Profanity Detection Library with scikit-learn. https://victorzhou.com/blog/better-profanity-detection-with-scikit-learn/

A Skill Resource Files Analysis

This section presents analysis of the files retrieved during skill interactions. Table 3 gives the exact breakdown of files.

Table 3: File Extensions per Web Resource

File Type	S3	CloudFront	Other	Total
aac	0	0	41	41
ashx	0	0	141	141
css	0	3	1	4
gif	5	0	0	5
jazz	0	0	1	1
jpeg	9	1	20	30
jpg	362	91	1002	1455
js	0	6	12	18
klassik	0	0	1	1
m3u	0	0	13	13
m3u8	0	429	118	547
m4a	1	0	9	10
mp3	494	210	1161	1865
mp4	2	0	4	6
opus	0	0	30	30
pls	2	0	33	35
png	584	174	1184	1942
svg	0	2	0	2
unknown	88	5	743	836
wav	0	0	1	1
Total	1547	921	4514	6982

The S3 hosted files were mainly image and audio files, with 61% image files and 32% audio files. CloudFront buckets contained similar content, with the ratios swapped: 70% audio and 29% image. Of interest are 3 CSS files served and 6 JavaScript files served. Otherwise, the resources served were split mainly between image and audio files, at 49% and 33%, respectively. Web (non-bucket) resources followed a similar distribution to S3 buckets, with slightly less than twice as many image files as audio files. Of note are .aac files, which are used for streamed audio from radio stations, and 12 JavaScript files. Overall, we retrieved 6,982 unique files across 10,407 skills.

B Crawler Skill Data


```
"skdetail": "Skill Details\nThis skill contains dynamic content.\nInvocation
Name: ego booster",
"des": "Description\nWhenever you're feeling small and unimportant, this skill
can help you out by lifting you up and telling you something positive about
yourself. Try it and see what effect it may have on you.",
"1": "\u201cAlexa Ask Ego Booster to tell me how he feels about me\u201d",
"cusper": "N",
"name": "EgoBooster",
"review": [],
"apprate": "4 out of 5",
"publ": "by Fabian Schneider",
"url": "https://www.amazon.com/Fabian-Schneider-Ego-Booster/dp/B06XWT6L9X/
ref=sr_1_2038?dchild=1&qid=1610883649&s=digital-skills&sr=1-2038",
"reviewnum": "1",
"2": "\u201cAlexa Ask Ego Booster to boost my ego\u201d",
"0": "\u201cAlexa Ask Ego Booster to give me a boost\u201d"
```


Figure 6: Skill information for the Ego Booster skill. Important information gathered includes ratings, reviews, developer, and invocations.

We gathered data from each skill page, including skill and developer name, rating, number of reviews, invocations, skill details, and presence of a developer policy or a security policy. Using this information, we gathered the top 25% skills from each category, as well as generated invocations and prompts for each skill. Figure 6 shows an example from the Ego Booster skill.