Your Echos are Heard: Tracking, Profiling, and Ad Targeting in the Amazon Smart Speaker Ecosystem

Umar Iqbal[•] Pouneh Nikkhah Bahrami[†] Rahmadi Trimananda[‡] Hao Cui[‡] Alexander Gamero-Garrido[¶]
Daniel Dubois[¶] David Choffnes[¶] Athina Markopoulou[‡] Franziska Roesner[•] Zubair Shafiq[†]

[•]University of Washington [†]University of California-Davis [‡]University of California-Irvine
[¶]Northeastern University

We answer frequently asked questions about the paper on alexaechos.com

Abstract

Smart speakers collect voice input that can be used to infer sensitive information about users. Given a number of egregious privacy breaches, there is a clear unmet need for greater transparency and control over data collection, sharing, and use by smart speaker platforms as well as third party skills supported on them. To bridge the gap, we build an auditing framework that leverages online advertising to measure data collection, its usage, and its sharing by the smart speaker platforms. We evaluate our framework on the Amazon smart speaker ecosystem. Our results show that Amazon and third parties (including advertising and tracking services) collect smart speaker interaction data. We find that Amazon processes smart speaker interaction data to infer user interests and uses those inferences to serve targeted ads on-platform (Echo devices) as well as off-platform (web). Smart speaker interaction leads to as much as 30× higher ad bids from advertisers. Finally, we find that Amazon's and skills' operational practices are often not clearly disclosed in their privacy policies.

1 Introduction

The convenience of voice input has contributed to the rising popularity of smart speakers [55], such as Amazon Echo [54], but it has also introduced several unique privacy threats. Many of these privacy issues stem from the fact that smart speakers record audio from their environment and potentially share this data with other parties over the Internet—even when they should not [62]. For example, smart speaker vendors or third-parties may infer users' sensitive physical (e.g., age, health) and psychological (e.g., mood, confidence) traits from their voice [84]. In addition, the set of questions and commands issued to a smart speaker can reveal sensitive information about users' states of mind, interests, and concerns. Despite the significant potential for privacy harms, users have little-to-no visibility into what information is captured by smart speakers, how it is shared with other parties, or how it is used by such parties.

Prior work provides ample evidence to support the need for greater transparency into smart speaker data collection, sharing, and use. For instance, smart speaker platforms have been known to host malicious third-party apps [59, 89], record users' private conversations without their knowledge [65, 66], and share users' conversations with strangers [83]. Further, platforms have patented several privacy-infringing practices to monetize voice input. For example, Amazon has a patent for advertising products to users based on inferences from physical and emotional characteristics of users' voices, e.g., targeting cough-drop ads at users with colds [71].

There is a clear need for auditing how smart speaker ecosystems handle data from their users' interactions. To facilitate such independent, repeatable audits, we need an approach that can work on unmodified, off-the-shelf devices, and that does not rely on disclosures provided by the smart speaker manufacturer. Conducting such an audit, however, requires addressing two key open challenges. First, commercially available smart speakers are black-box devices without open interfaces that allow independent researchers to expose what data is collected or how they are shared and used. Second, when data gathered from a smart speaker is sent over the Internet, there is no way to isolate how the data is further shared and used.

In this paper, we address these challenges by building an auditing framework that measures the collection, usage, and sharing of smart speaker interaction data. *Our key insight is that data collection and sharing over the Internet can be inferred through its usage in targeted advertisements.* Namely, we can create multiple personas with different smart-speaker usage profiles, and test whether those personas receive statistically significantly different advertisements and bid values. This, in turn, allows us to infer how data was shared and used.

To evaluate the effectiveness of this approach, we focus on Amazon's smart speaker platform, as it is the largest platform (46 million devices in the US [47] and 200K third-party applications [63]). To address the first challenge, we set up a custom Raspberry RPi (RPi) router [46] that allows us to capture the endpoints contacted by commercial Amazon Echos. We also emulate an Amazon Echo by instrumenting Alexa

Voice Service (AVS) SDK [16] (referred to as AVS Echo) to capture the collected data because there is no trivial way to decrypt TLS traffic from commercial Amazon Echos,

To address the second challenge, we conduct controlled experiments where we intentionally expose data with voice commands to an Amazon Echo and look for its usage on-platform (i.e., on an Amazon Echo) and off-platform, (i.e., on the web). We expose data by installing and interacting with apps (called *skills* in the Amazon Echo ecosystem) from different categories according to *personas* that represent users with different interests, e.g., a "fashion" persona is configured to install and interact with skills from the fashion category.

To determine whether our personas' smart speaker interactions are used or shared, we look for evidence in online targeted advertising [58, 61, 79]. We measure targeting across two modalities and multiple devices: audio ads served by Amazon Echos and display ads served by websites. By comparing ad content and ad auction bid values across personas and carefully controlling what information is exposed to other parties, we can identify when smart speaker interactions are likely the cause of ad targeting, and thus infer that data was shared and/or used for that purpose.

Key contributions. Our auditing framework allows us to answer three crucial questions:

1. *Which organizations collect and propagate user data?* Amazon Echo interaction data is collected by both Amazon and third-parties, including advertising and tracking services. As many as 41 advertisers sync their cookies (i.e., share data) with Amazon. These advertisers further sync their cookies with 247 other third parties, including advertising services.
2. *Is smart speaker interaction data used by either Amazon or third-party apps beyond purely functional purposes, such as for targeted advertising?* Amazon processes Echo interactions to infer user interests. Our measurements indicate the usage of inferred interests for on-platform (i.e., audio ads), off-platform (i.e., web ads), and cross-device (i.e., non-Echo device) ad targeting. Advertisers bid as much as 30× higher on some personas. It is unclear if third-party skills infer user interests and target personalized ads.
3. *Are data collection, usage and sharing practices consistent with the official privacy policies of Amazon and third-party skills?* Our measurements indicate that Amazon's and skills' operational practices are often not clearly disclosed in their policies or other claims. For example, Amazon's inference of advertising interests from users' Echo interactions seems to be inconsistent with their public statements [77, 85]. Similarly, over 70% skills do not even mention Alexa or Amazon and only 10 (2.2%) skills are clear about data collection practices in their privacy policies.

In summary, we find strong evidence that smart speaker interactions are used for the purpose of targeting ads, and that this ad targeting implies significant data sharing across multiple parties. To further strengthen and enable new forms of auditing, we argue that substantial additional transparency is needed in the smart speaker ecosystem. To that end, we will make our code and data publicly available upon publication.

2 Background & Motivation

2.1 Amazon Echo & Alexa

In this paper, we study Amazon's smart speaker platform, the most widely used platform with more than 46 million devices in the US [47]. Amazon's smart speakers are called Echo and they are powered by Alexa. Alexa is a voice assistant that responds to user requests conveyed through voice input. Although Alexa can respond to a wide variety of general-purpose requests, it is not well-suited for specialized tasks, e.g., ordering a pizza from a particular restaurant. Thus, to augment Alexa, Amazon allows third party services to build and publish applications called *skills* on the Alexa marketplace. As of 2020, the Alexa marketplace hosts more than 200K third party skills [63].

2.2 Privacy issues

The inclusion of third party skills poses a privacy risk to the users of Amazon Echo. Accordingly, Amazon imposes a set of platform policies to mitigate potential privacy risks of third party skills. Amazon restricts skills from collecting sensitive information, e.g., social security and bank account numbers [7, 8], and requires user permission to allow access to personal information, e.g., email, phone, location [19]. To enforce the aforementioned policies, Amazon has a skill certification process that aims to filter malicious skills before they can be published on the marketplace [6]. However, prior research has shown that policy-violating skills can get certified [59] and thousands of skills on the Alexa marketplace violate platform policies [89].

Smart speakers also handle particularly sensitive voice input. Voice input in its raw form, i.e., voice recording, can be used to infer several physical (e.g., age, health) and psychological characteristics (e.g., mood, confidence) of the user [84]. Voice input in its processed form, i.e., transcript of voice recording, can leak sensitive information (e.g., private conversations) about the user. Even the information that results in execution of a voice input/command (e.g., a user querying a third party skill about a medical condition) can leak sensitive information about the user. Amazon aims to limit some of these privacy issues through its platform design choices [5]. For example, to avoid snooping on sensitive conversations, voice input is only recorded when a user utters the *wake word*, e.g., Alexa. Further, only keywords from transcripts of voice input are shared with third party skills, instead of the raw audio [35]. However, despite these design choices, prior research has also shown that smart speakers often *misactivate* and unintentionally record conversations [62]. In fact, there have been real-world instances where smart speakers recorded user conversations, without users ever uttering the wake word [66].

Smart speakers also store the *voice data*, i.e., voice recordings, their transcripts, and the information about the resulting action. This raises two privacy concerns. First, smart speaker vendors can use voice data for targeting ads as proposed in an Amazon patent [71] and as recently admitted by Amazon in response to our research (see Appendix 11.1 for details). Second, voice data may be shared with other parties. For example, when a user interacts with a third party skill, the keywords from the transcriptions of voice input are shared with the third party. In these cases, neither users nor Amazon have any visibility or control on the processing, sharing, and selling of voice data. Third party skills often do not publish their privacy policies, nor adhere to them even when they do [63].

2.3 Proposed auditing framework

To the best of our knowledge, prior work lacks an in-depth analysis of the collection, usage, and sharing of voice data in smart speaker ecosystems. To fill this gap, we systematically analyze the data collection, usage, and sharing practices of Amazon's smart speaker platform including third party skills—the most widely used platform. We conduct controlled experiments where we intentionally leak user's voice data and observe platform's behavior from three perspectives: *(i) network traffic* exchanged by smart speakers, *(ii) advertisements* served to smart speaker users, and *(iii) privacy policies* published by third-party skills. Our goal is to combine these perspectives to answer the following research questions.

RQ1: Which organizations collect and propagate user data? We use network traffic flows (e.g., remote endpoints) to measure data collection and sharing by Amazon and third party skills. While we are able to intercept and observe communication between Amazon and some third parties, we otherwise find that the Amazon ecosystem uses opaque practices and does not offer any debugging interfaces which can be leveraged to analyze data collection, usage, and sharing.

RQ2: Is voice data used by either Amazon or third-party apps beyond purely functional purposes, such as for targeted advertising? We measure advertisements to infer data usage and sharing by Amazon and third-party skills. To this end, we focus on detecting behaviorally targeted web and audio ads. We study targeting in web ads because web publishers almost universally employ well-established programmatic advertising protocols [30, 42]. We also study targeting in audio ads even though smart speaker advertising ecosystem is relatively nascent.[1]

RQ3: Are data usage and sharing practices compliant with privacy policies? We extract key elements from privacy policies of Amazon and third party skills. We compare privacy policies with our network traffic measurements to assess the compliance of data collection, usage, and sharing practices.

[1]Amazon only allows audio ads on streaming skills [3] and typically requires rather high minimum ad spend commitment from advertisers [13].

2.4 Distinction between voice recordings, transcripts, and the resulting actions

It is noteworthy that our framework cannot pinpoint whether the voice recordings, their transcript, or the information about the resulting action is abused. Mainly because there is no trivial way to leak one without leaking the other. Specifically, leaking voice input/recording results in automatic leakage of transcripts and the information about the resulting action. Similarly, neither the transcripts nor the information about the resulting action can be leaked without leaking the voice input/recording. Most importantly, users may not find any meaningful distinction between the leakage of privacy through voice recording, transcripts, and the resulting actions (as we later show in Section 9). In this paper we refer to the voice recordings, their transcripts, and the information about the resulting action, collectively as *voice data*.

3 Measuring Tracking, Profiling, Ad Targeting

In this section, we describe our methodology to measure tracking, profiling, and ad targeting by Amazon and third-party skills. Figure 1 presents the overview of our approach. At a high level, we first intentionally leak data by interacting with skills on Amazon Echo, then measure data tracking by intercepting network traffic, profiling by requesting data from Amazon, and ad targeting by analyzing ads on popular websites and music streaming skills.

3.1 Leaking data

3.1.1 Simulating interest personas

We simulate nine interest (treatment) personas by installing and interacting with skills from nine different Alexa marketplace categories: Connected Car, Dating, Fashion & Style, Pets & Animals, Religion & Spirituality, Smart Home, Wine & Beverages, Health & Fitness, and Navigation & Trip Planners. We simulate several personas because the nature of tracking, profiling, and ad targeting might differ across different skill categories. Each interest persona is referred by the respective skill category name, and collectively as *interest personas*.

Skill installation. As a first step, we create dedicated Amazon accounts for each persona and use them to configure Amazon Echos (4th generation Amazon Echo smart speakers). To avoid contamination across personas, we configure each Amazon Echo through a fresh browser profile and assign unique IP address to each device (all IPs geolocate to the same location). We then use a Selenium [45] based web crawler to programmatically visit the Alexa skill marketplace, and iteratively install and enable the top-50 skills (based on the number of reviews) for each category. We use the dataset

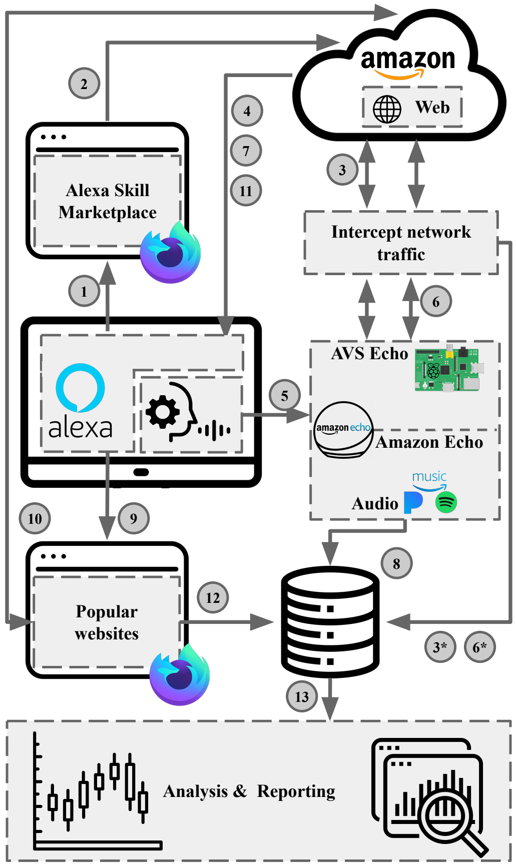


Figure 1: Approach overview: We link Amazon Echo to the Alexa web companion app and visit Alexa skill marketplace to install skills (1–4), we then interact with the installed skills by uttering sample invocation utterances listed in skill description (5–8), we then visit popular websites while logged into Amazon account and Alexa web companion app (9–11). We record network traffic to/from Amazon Echo and AVS Echo while installing and interacting with the skills (3^* and 6^*). We record audio ads from music streaming skills (8). We record web ads on popular websites (12). After recording the data we analyze it to measure tracking, profiling, and ad targeting and its compliance with privacy policies (13).

released in [63] to extract top skills from each category. If prompted, we enable all of the requested permissions by a skill. It is noteworthy that we do not link accounts for skills that require to link an account. Our rationale for this methodological choice is to sidestep the non-trivial account linking process, that typically requires creating an account for the online service and linking a physical IoT device, e.g., *iRobot* skill requires to link a robot vacuum cleaner with the skill [70].

Skill interaction. After installing each skill, we interact with it by programmatically uttering sample invocations listed by each skill. We also parse skill descriptions to extract additional invocation utterances provided by the skill developer. We interact with the Amazon Echo by iteratively uttering each skill's invocations. In case Alexa expects a follow up response[2] or has a response of more than 30 seconds, e.g., playing a song, we terminate the interaction by uttering *Alexa, Stop!*. Note that a minute chunk of generic utterances, such as *Alexa, give me hosting tips*, were redirected to Alexa instead of the skills. We surmise that it could be because of the unavailability of the skill's backend server at the time of interaction, a bug in the skill, or an unexpected sample utterance listed by the skill developer.

3.1.2 Simulating control personas

In addition to the nine interest (treatment) personas, we also simulate four control personas. One control persona is linked to an Amazon account and an Amazon Echo and referred to as *vanilla* persona. The remaining three personas are primed by iteratively visiting top-50 websites from health, science, and computer categories [9], and are referred to as *web health*, *web science*, and *web computer* personas. We use OpenWPM [64], an open-source web measurement tool to prime web personas. Similar to interest personas, to avoid contamination across control personas, we configure each of them through a fresh browser profile and assign unique IP address to each persona.

Control personas serve as a baseline and allow to associate deviation to the treatment applied to the interest persona in question. Vanilla persona serves as a baseline for tracking and profiling the information that the user is an Amazon consumer and owns an Amazon Echo. Web health, science, and computer personas serve as a baseline for standard data tracking and profiling on the web, about users with respective interests. We pick web health, science, and computer personas because prior work has shown these personas to attract most targeting [61]. The additional comparison with web personas allow us to better contextualize the results, because as compared to smart speakers, ad targeting has been extensively studied on the web [61, 78, 79].

3.2 Capturing network traffic

We capture outgoing and incoming network traffic, to and from, Amazon Echos to measure data collection by Amazon and skills during skill installation and interaction. Since, Amazon Echo does not provide any interface to monitor network traffic on the device, we intercept network traffic on the router. To this end, we set up a custom Raspberry Pi (RPi) based router [46] to intercept incoming and outgoing network traffic. For each skill, we enable `tcpdump` on the RPi router, install the skill, interact with the skill, uninstall the skill, and disable `tcpdump`. Enabling and disabling `tcpdump` allow us to cleanly associate network traffic to each skill. Similarly,

[2] If Alexa expects a follow up response and the response is not provided, Alexa asks for the response a few times. We match the last two responses to determine if a follow up response is expected.

uninstalling each skill before installing the next one ensures that we associate the correct network traffic to each skill.

Unencrypted network traffic. Since we can only capture encrypted network traffic on the router, we lack visibility on the collected data. To enhance our coverage, we additionally simulate an Echo device by instrumenting Alexa Voice Service (AVS) SDK [16] and running it on a Raspberry Pi (RPi)—we call it AVS Echo. We use the instrumented AVS Echo to intercept and log the payload of each packet before it is encrypted and sent over the network. The network traffic captured through the AVS Echo allows us to examine all the data, including any personally identifiable information (PII), sent in the network traffic, which otherwise is not possible to observe in the encrypted network traffic captured from the Amazon Echo on the RPi router. However, it is important to note that skills that stream content, e.g., music, podcast, are not supported on un-certified Alexa on AVS Echo [44]. Further, unlike commercial Amazon Echos that can communicate with Amazon and third-party endpoints, AVS Echo only communicates with Amazon.

Inferring origin. Both encrypted (captured on commercial Echo devices) and unencrypted (captured on instrumented AVS Echo) network traffic contain the IP addresses of contacted endpoints. We resolve IP addresses to domain names by using the information from Domain Name System (DNS) packets in network traffic. We further map domain names to their parent organization by leveraging information from DuckDuckGo [24], Crunchbase [21], and WHOIS.

3.3 Capturing advertisements

We rely on ad content and advertisers' bidding behavior to infer data usage and sharing. Ad content can reveal the ad topic and consequently the user interests that advertisers might have inferred from the leaked Amazon Echo interaction data. However, ad content may lack objective or discernible association with the leaked data. For example, active advertising campaigns may lack apparent association with the leaked data or advertising models may interpret user interests differently. We try to offset subjectivity by also relying on advertisers' bidding behavior to infer the usage and sharing of smart speaker interaction data. Prior research [61, 78, 79] has shown that the advertisers' bidding behavior is influenced by their preexisting knowledge of the users, which typically results in high bid values. Thus, if we encounter high bid values from advertisers, a likely cause is the usage and sharing of Amazon Echo interaction data.

Web advertisements. Since header bidding protocol [30] allows to observe bid values at the client side, we collect ad bids and ad images on header bidding supported websites both after skill installation and skill interaction. To this end, we first identify top websites that support `prebid.js` [40], the most widely used implementation of header biding protocol [31], and then visit those websites to capture bids and ad images. We extend OpenWPM [64] to identify and capture data on `prebid.js` supported websites. To identify `prebid.js` supported websites, we crawl Tranco top websites list [72] and probe for `prebid.js` version, through an injected script that calls `pbjs.version`. We treat a website as prebid supported, if we receive a non-null `prebid.js` version. We stop the crawl as soon as we identify 200 prebid supported websites. We then crawl the `prebid.js` supported websites and intercept bidding requests. Specifically, we inject a script on the webpage and collect the bids by calling `pbjs.getBidResponses` function. In case the website has not received any bids, we request the bids ourselves by calling `pbjs.requestBids` function.

In order to more accurately simulate user behavior, we enable OpenWPM's bot mitigation and wait for 10–30 seconds between webpage visits. It is important to note that we crawl the `prebid.js` supported websites using the same browser profiles, that are logged into Amazon account and Alexa web companion app, and IP addresses used to configure interest and vanilla personas (Section 3.1). The browser profiles and IP addresses connect personas with browsers and allow us to collect the advertisements targeted to the personas.

Interpreting bids. In addition to user interests, advertisers consider several factors, e.g., day of the week, website popularity, to determine the bid values [78, 79]. We try to minimize the variability by keeping conditions consistent across personas. Specifically, we use identical hardware/software, collect bids at the same time (simultaneously), from the same location, and on the same websites, for all personas. In addition, we only consider bids from ad slots that are successfully loaded across all personas, because bid values vary by ad slots [79] and advertisers may not bid on all ad slots across all personas. We relatively compare bid values across control and interest (treatment) personas because the absolute values can change over time, e.g., travel advertisements may get higher bids around holidays. Since it is non-trivial to reverse engineer and control for all the factors incorporated by advertisers, we crawl and extract bids from the `prebid.js` supported websites several times, i.e., 6 times before interacting with skills and 25 times after interacting with skills[3], to further minimize the variability in bid values.

Capturing requests/responses. In addition to collecting ad bids and images, we also record the network requests and responses while crawling popular websites. Network traffic allows us to measure data sharing, e.g., cookie syncing [67], between Amazon and its advertising partners. Note that the traffic captured while crawling is different from the traffic captured from Amazon Echos and AVS Echos (Section 3.2).

Audio advertisements. Considering the rapid growth of audio advertising, we also try to infer data usage and sharing

[3]We terminated the experiment after 6 iterations after skill installation because we did not notice any personalization. We continued to crawl 25 times after skill interaction because we noticed personalization (Section 6) and wanted to minimize the variability in bid values.

through audio ads, despite their shortcomings (mentioned in Section 2). We capture audio ads played on Amazon Music [10], Spotify [49], and Pandora [36]. We include Amazon Music to determine if Amazon (the platform operator) personalizes audio ads, while the other two are popular streaming services [15, 50] with over 10,000 reviews on the Amazon Alexa skill marketplace [36, 49]. We programmatically issue a command to Amazon Echo, i.e., *Alexa, play today's top hits on Amazon music / Spotify / Pandora*, to play music on each platform and record the ads played in-between songs. Since ads are played at variable intervals in-between songs, we stream music for 6 hours on each skill to collect a significant number of ads. We then automatically transcribe the recorded audio files [80] and manually extract ads from transcripts.

It is noteworthy that we only capture audio ads on two interest personas (Connected Car, Fashion & Style) where we expect most personalization (see Section 6.7), and the Vanilla persona for baseline comparison. We reduce the number of personas compared to our web experiments because of the time- and labor-intensive nature of collecting and processing audio ads. Specifically, to capture audio ads, we place Amazon Echos in insulated environments to avoid interference; a human coder then manually inspects both the audio recording and their transcripts to identify ads (rather than song lyrics). We place Amazon Echos in 3 different rooms, one for each persona, and collect audio ads simultaneously to reduce variability. We then manually identify ads from 54 hours (3 personas $\times$ 3 skills $\times$ 6 hours) of audio transcripts.

4 Network Traffic Analysis

In this section we analyze network traffic to identify online services that collect user data and investigate the functionality offered by these services.

4.1 Amazon has the best vantage point to track user activity

We first analyze network traffic to identify the services that collect user data and the type of data collected by them. Table 2 presents the list of domains contacted by skills. Unlike more established platforms, e.g., mobile, the majority of the traffic in Amazon's smart speaker ecosystem goes to the the device manufacturer. Specifically, we note that, 446 (99.11%), 2 (0.45%), and 31 (6.89%) of the skills contact domains that belong to Amazon, skill vendors, and third parties, respectively. Four (0.89%) skills failed to load. All active skills contact Amazon because Amazon mediates communication between skills and users, i.e., Amazon first interprets the voice input and then shares it with the skill [35]. Another possible explanation for a large number of network traffic flows to Amazon could be the hosting of skills on Amazon's platform [4]. *Garmin* [27] and *YouVersion Bible* [53] are the only skills that send traffic to their own domains.

Figure 2 shows the network flows from skills to domains, their functionality, and their parent organizations. Corroborating with the results from Table 2, we note that most network flows involve Amazon. We also note that the skills in most categories, except for Smart Home, Wine & Beverages, Navigation & Trip Planners, contact third party services.

Table 1 presents the details of data shared by skills (based on unencrypted network traffic captured on AVS Echo). As expected, voice recording is collected when a skill is installed and enabled. Further, 326 (72.44%) skills collect persistent identifiers, namely user and skill IDs[4], 434 (96.44%) collect user preferences, and 385 (85.55%) collect device events. We also note that 8.59% of the skills that collect persistent identifiers also send data to third-party domains.

4.2 Data is leaked to advertisers and trackers

We next analyze the functionality offered by services that collect user data. Several domains contacted by skills offer audio advertising and tracking services (rows highlighted in gray in Table 2). We rely on filter lists [37] and manual investigations to detect advertising and tracking services. Table 3 provides the distribution of functional and advertising domains contacted by skills. We note that 9.4% of all network traffic, including 1.5% third-party network traffic, supports advertising and tracking functionality. We note that `device-metrics-us-2.amazon.com`, used by Amazon to collect device metrics [57], is the most prominent tracking domain. Most contacted third-party advertising and tracking services include Megaphone (`megaphone.fm`) and Podtrac (`podtrac.com`), both of which specialize in audio advertising and tracking services. We note that prominent skills, such as *Genesis* [28] and *Men's Finest Daily Fashion Tip* [34] with 398 and 13 reviews, contact third-party advertising and tracking services. Despite Amazon's Alexa advertising policy restricting non-streaming skills from playing ads [3, 38], we find that six non-streaming skills contact third-party advertising and tracking services. Surprisingly, we note that these skills do not play any advertisements, despite including advertising services. It is unclear as to why non-streaming skills include advertising and tracking services and if such skills should be flagged during skill certification [14].

Table 4 and 5 further provide the distribution of advertising and tracking domains by personas and skills. From Table 4, we note that skills in five personas contact third-party advertising and tracking services, where skills in Fashion & Style persona contact the most advertising and tracking services. From Table 5, we note that skills contact several advertising and tracking services. The skill *Garmin* [27] even contacts as much as 4 advertising and tracking services.

[4] These IDs persist across sessions but are unique across skills [20, 23].

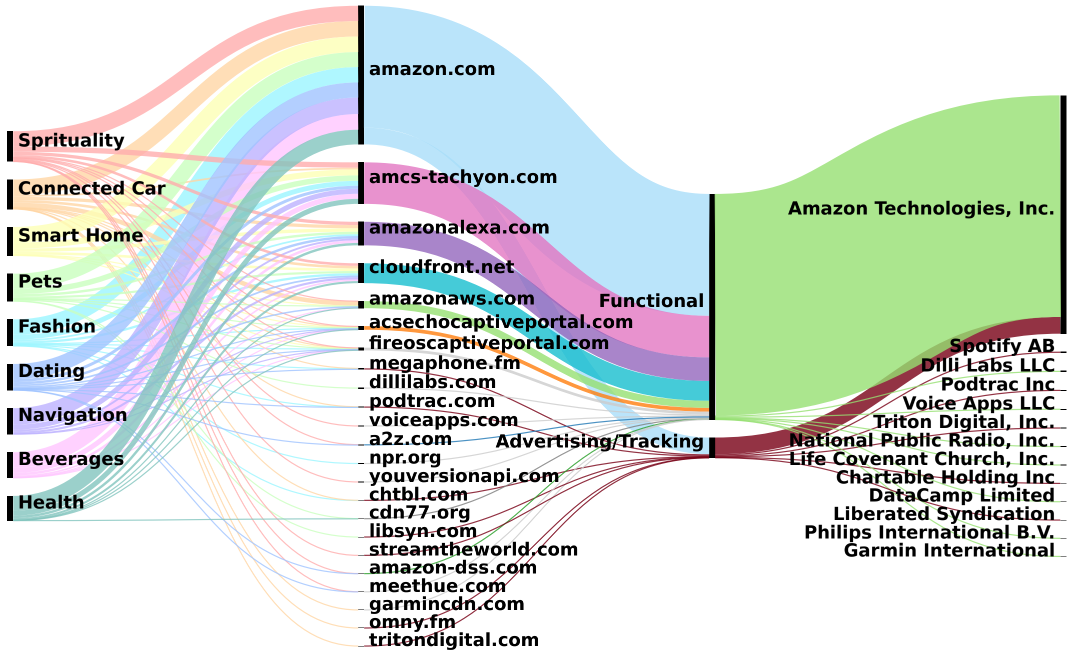


Figure 2: Network traffic distribution by persona, domain name, purpose, and organization.

5 Data Profiling Analysis

In this section, we analyze whether our data leakage results in profiling of personas by Amazon. Since Amazon allows users to access data collected about them, we request data for interest and vanilla personas [11]. The data contains detailed information about device diagnostics, search history, retail interactions, Alexa, advertising, and other Amazon services. We are mostly interested in advertising interests inferred by Amazon based on skill installation and interactions. We request data thrice, once after skill installation and twice after skill interaction, to see the evolution in inferred interests over time. Since advertising interests are inferred instantly and made available to users for download within 2 days [90], we request user interests after 3 days of skill installation and 8 and 31 days of skill interaction. Amazon on average took around 12 days to return the inferred interests after our request.

Table 6 presents the advertising interests inferred by Amazon for Health & Fitness, Fashion & Style, and Smart Home personas. For remaining personas, Amazon did not return any interests. We note that both skill installation and interaction leads to interests inference by Amazon. With only skill installation, Amazon infers that Health & Fitness persona is interested in *Electronics* and *DIY & Tools*. Skill interaction, further allows Amazon to infer interests for Fashion & Style and Smart Home personas and also refine interests for Health & Fitness persona. Table 6 shows that some of the interests even have discernable relevance to the personas. For example, *Fashion* and *Beauty & Personal Care* interests have discernable relevance with Fashion & Style persona and *Home & Kitchen* interests have discernable relevance with Smart Home persona. It is noteworthy that for our second data request after interaction, Amazon did not return advertising interest files for Health & Fitness, Wine & Beverages, Religion & Spirituality, Dating, and vanilla personas. To eliminate a one-off technical issue, that may have resulted in absence of advertising interest files, we again requested data from Amazon but the advertising interest files were still absent. Though the exact reason behind the absence of files is unclear, Amazon cannot be reliably trusted to provide transparency in data usage.

It is notable that advertising interest inference can be interpreted as inconsistent with Amazon's public statements and policies. Specifically, in a statement, Amazon mentioned that they do "not use voice recordings to target ads" [77, 85]. Amazon's privacy policy neither explicitly acknowledges nor denies the usage of Echo interactions for ad targeting [12]. Amazon Alexa Privacy Hub, a dedicated webpage that explains privacy practices of Alexa, only states that the Alexa interactions are used for personalizing user experience (e.g. improve speech recognition) and to build a more inclusive Alexa (e.g., understand different accents) [5]. Similarly, Alexa Device FAQs, that answers 100+ frequently asked questions about Alexa, also do not mentions the usage of Echo interactions for ad targeting [2].

Our results suggest that Amazon processes voice recordings to infer user interests for ad targeting. The potential inconsistency between policies/statements and actual practices raises questions about Amazon's commitment to only using user data for stated purposes.

Category	Data Type(s)	Skill Disclosures				Example terms in privacy policies	
		Clr.	Vag.	Omi.	No Pol.	Amazon	Skills
Voice inputs	voice recording	20	18	147	258	voice recording	audio recording , sensory info
Persistent IDs	customer / user ID	11	9	38	84	unique identifier	anonymized ID , UUID
	skill ID	0	11	85	230	cookie	
User preferences	language	0	3	5	10	time zone setting ,	regional and language settings ,
	timezone	0	3	5	10	settings preferences	app settings
	other preferences	0	40	139	255		
Device events	audio player events	0	60	99	226	device metrics , Amazon Services metrics	usage data , interaction data

Table 1: Data type analysis results. “Skill Disclosures” column presents the numbers of skills that have clear , vague and omitted disclosures for a certain “Data Type”, and number of skills with no policy .

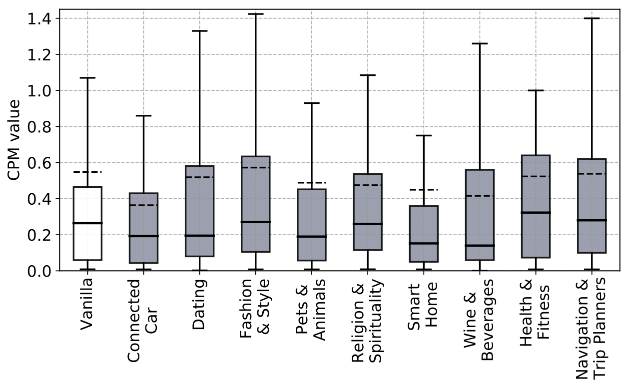


Figure 3: CPM values across vanilla (control) and interest (treatment) personas on common ad slots with skill installation. Solid and dotted lines in bars represent median and mean, respectively. Vanilla persona does not involve skill installation.

6 Ad Targeting Analysis

In this section, we analyze whether Amazon uses the interests inferred from Echo interactions for ad targeting.

6.1 Skill installation without user interaction does not show apparent targeting

We first analyze advertisers bidding behavior for vanilla and interest personas without any user interaction to evaluate if only skill installation leads to personalized ad targeting. Figure 3 presents bid (CPM)[5] values across vanilla and interest personas on common ad slots without user interaction. It can be seen from Figure that without user interaction, there is no discernible difference between vanilla and interest personas. Only Fashion & Style personas has slightly higher, statistically insignificant, mean bid value than vanilla persona.

6.2 User interaction leads to higher bid values

Next, we analyze advertisers bidding behavior for vanilla and interest personas with user interaction to evaluate if interaction with skills leads to personalized ad targeting. Figure 4 presents bid (CPM) values across vanilla and interest personas on common ad slots with user interaction. In contrast to bid values without user interaction (Figure 3), with user interaction (Figure 4) the bid values are significantly higher for interest personas as compared to vanilla persona. Table 8 shows the median and mean values for interest and vanilla personas with user interaction. It can be seen from the table that median bids for all interest personas, except for Health & Fitness, are 2× higher than vanilla persona. Similarly, mean bids for four interest personas, i.e., Fashion & Style, Religion & Spirituality, Wine & Beverages, and Health & Fitness, are 2× higher than vanilla persona. We note that the bid values for Health & Fitness and Fashion & Style go as much as 30× and 27× higher than the mean of vanilla persona.

6.3 High absolute bid values with only skill installation

We note that the absolute bid values for vanilla and interest personas with only skill installation are higher than that of vanilla and interest personas with user interaction. While it is impossible to know the exact reason for high bid values for personas with only skill installation, a possible explanation could be our data collection during the holiday season, i.e., around Christmas 2021.[6] To rule out the impact of holiday season, we compare the bids values with only skill installation

[5]CPM (cost per mille) is the amount an advertiser pays a website per thousand visitors who see its advertisements. Bids are expressed in CPM.

[6]We collected advertisements (including bids) both before and after interacting with the skills from 12/08/21 to 12/22/21 and 12/28/21 to 02/04/22, respectively. We installed the skills on 12/07/21 and we interacted with the skills on 12/27/21.

Org.	Domains	Skills
Amazon	*(11).amazon.com	895
	prod.amcs-tachyon.com	305
	api.amazonalexa.com	173
	*(7).cloudfront.net	146
	device-metrics-us-2.amazon.com	123
	*(4).amazonaws.com	52
	acsechocaptiveportal.com	27
	fireoscaptiveportal.com	20
	ingestion.us-east-1.prod.arteries.alexa.a2z.com	7
	ffs-provisioner-config.amazon-dss.com	2
Skill vendor	*(2).youversionapi.com	2
	static.garmincdn.com	1
Third party	dillilabs.com	9
	*(2).megaphone.fm	9
	cdn2.voiceapps.com	7
	*(2).podtrac.com	7
	*(2).pod.npr.org	4
	chtbl.com	3
	1432239411.rsc.cdn77.org	3
	*(2).libsyn.com	3
	*(3).streamtheworld.com	3
	discovery.meethue.com	2
	turnernetworksales.mc.tritondigital.com	1
	traffic.omny.fm	1

Table 2: Amazon, skill vendors, and third-party domains contacted by skills. "Org." column refers to organization. "Skills" column represents the count of skills. Advertising and tracking domains are shaded with grey. Subdomains counts are represented with *(#), e.g., *(11).amazon.com represents requests to 11 subdomains of amazon.com.

and with skill interaction that were collected close to each other. Specifically, we compare the bids from last three iteration of without interaction with bids from first three iterations of with interaction, that were crawled within a close time span.[7] Table 7 presents mean bid values without and with user interaction. It can be seen that the interest personas with interaction receive higher bids than control persona. Whereas no discernible differences exist for without interaction configurations.

Although the timing affects the bid values, we believe that it does not impact our findings. Specifically, our objective is to measure the effect of treatment, i.e., skill installation or interaction, on interest (treatment) personas as compared to the vanilla (control) persona. The relative comparison of bid values between control and interest (treatment) personas suffices to measure the effect of treatment. It means that, if we see statistically significant differences in bid values between control and interest (treatment) personas, we can confidently attribute the differences to the applied treatment, i.e., skill installation or interaction.

[7] From 12/20/21 to 12/22/21 for skill installation and from 12/28/21 to 01/12/22 for skill interaction.

Organization	Functional	Advertising & Tracking	Total
Amazon	90.04%	6.80%	96.84%
Skill vendor	0.17%	0%	0.17%
Third party	1.49%	1.50%	2.99%
Total	91.70%	8.3%	100%

Table 3: Distribution of advertising / tracking and functional network traffic by organization.

Persona	Advertising & Tracking	Functional
Fashion & Style	9	4
Connected Car	7	0
Pets & Animals	3	11
Religion & Sprituality	3	10
Dating	5	1
Health & Fitness	0	1

Table 4: Count of advertising/tracking and functional third-party domains contacted by personas.

6.4 After user interaction interest personas receive statistically higher bids than vanilla persona

We perform the Mann-Whitney U test to analyze whether interest personas after user interaction receive significantly higher bids than vanilla persona. Since we perform multiple comparisons, we adjust our statistical significance tests with the Holm-Bonferroni correction method. Our null hypothesis is that the bid distributions for interest personas are similar to vanilla persona. Our alternative hypothesis is that the bid distributions for interest personas are higher than the vanilla persona. We reject the null hypothesis when the p-value is less than 0.05. In addition to p-value, we also report the effect size (rank-biserial coefficient). Effect size ranges from -1 to 1, where -1, 0, and 1 indicate stochastic subservience, equality, and dominance of interest persona over vanilla persona. Effect size between 0.11–0.28, 0.28–0.43, and ≥ 0.43 are considered small, medium, and large, respectively.

Table 9 presents the results of statistical significance tests. We note that six interest personas have significantly higher bids than vanilla persona with medium effect size. For the remaining three interest personas, i.e., Smart Home, Wine & Beverages, and Health & Fitness, the differences are not statistically significant.

6.5 After user interaction interest personas are targeted personalized ads

Next, we analyze the ads delivered through `prebid.js` to personas after user interaction. In total, we receive 20,210 ads across 25 iterations. Since ads may lack any objective or even discernible association with the leaked interests, as discussed in Section 3.3, we resort to manual analysis of ads. However, manual ad analysis is a tedious task and it is not feasible to

Skill name	Advertising & Tracking
Garmin [27]	chtbl.com traffic.omny.fm dts.podtrac.com turnernetworksales.mc.tritondigital.com
Makeup of the Day [32]	*(2).megaphone.fm play.podtrac.com chtbl.com
Men's Finest Daily Fashion Tip [33]	play.podtrac.com *(2).megaphone.fm
Dating and Relationship Tips and advices [22]	play.podtrac.com *(2).megaphone.fm
Charles Stanley Radio [17]	*(2).streamtheworld.com

Table 5: Top-5 skills that contact third-party advertising and tracking services. Subdomains counts are represented with *(#), e.g., *(2).megaphone.fm represents two subdomains of `megaphone.fm`.

Config.	Persona	Amazon inferred interests
Installation	**Health & Fitness**	Electronics Home & Garden: DIY & Tools
Interaction (1)	**Health & Fitness**	Home & Garden: DIY & Tools
	Fashion & Style	Beauty & Personal Care Fashion Video Entertainment
	Smart Home	Electronics Home & Garden: DIY & Tools Home & Garden: Home & Kitchen
Interaction (2)	**Fashion & Style**	Fashion Video Entertainment
	Smart Home	Pet Supplies Home & Garden: DIY & Tools Home & Garden: Home & Kitchen

Table 6: Advertising interests inferred by Amazon for interest personas.

analyze thousands of ads. To this end, we only manually analyze ads from Amazon and ads from installed skill vendors in their respective personas (e.g., an ad from Ford in Connected Car persona because it contains the *FordPass* skill) because we expect these ads to be the most personalized. We consider an ad to be personalized if it is only present in one persona and references a product in the same industry as the installed skills (e.g., an ad for a vehicle is shown to the Connected Car persona). While any manual labeling process is subject to human error and subjectivity, we argue that our definition is sufficiently concrete to mitigate these concerns.

In total, we filter 79 ads from installed skills' vendors in their respective personas and 255 ads from Amazon ads for manual analysis. Out of the 79 ads from installed skills vendors, 60, 12, 1, and 1 are from Microsoft, SimpliSafe, Samsung, and LG in Smart Home persona, respectively. Out of

Persona	No Interaction	Interaction
Connected Car	0.364	0.311
Dating	0.519	0.297
Fashion & Style	0.572	0.404
Pets & Animals	0.492	0.373
Religion & Spirituality	0.477	0.231
Smart Home	0.452	0.349
Wine & Beverages	0.418	0.522
Health & Fitness	0.564	0.826
Navigation & Trip Planners	0.533	0.268
Vanilla	0.539	0.232

Table 7: Mean bid values without and with interaction across interest and vanilla personas that were collected close to each other.

Persona	Median	Mean
Connected Car	0.099	0.267
Dating	0.099	0.198
Fashion & Style	0.090	0.403
Pets & Animals	0.156	0.223
Religion & Spirituality	0.120	0.323
Smart Home	0.071	0.218
Wine & Beverages	0.065	0.313
Health & Fitness	0.057	0.310
Navigation & Trip Planners	0.099	0.255
Vanilla	0.030	0.153

Table 8: Median and mean bid values (CPM) for interest (treatment) and vanilla (control) personas with user interaction. Vanilla persona does not involve user interaction.

the remaining 5, 3 are from Ford and 2 are from Jeep in Connected Car persona. However, none of the ads from installed skills vendors are exclusive to the personas where their skills are installed, which indicates that these ads do not reveal obvious personalization.

Ads from Amazon do seem to be personalized to personas. Table 10 presents the unique from Amazon that show apparent personalization. Health & Fitness and Smart Home personas receive unique ads with apparent personalization, whereas Religion & Spirituality and Pets & Animals receive unique ads but without any apparent personalization. The dehumidifier ad (Figure 5a) appears to have an association with the *Air Quality Report* skill [1] and the essential oils ad appears to have an association with the *Essential Oil Benefits* skill [26] in Health & Fitness persona. The dehumidifier ad appeared 7 times across 5 iterations and the essential oils ad appeared once in Health & Fitness persona. The vacuum cleaner and vacuum cleaner accessories ads from Dyson appear to have an association with the *Dyson* skill [25]; both ads appeared once in Smart Home persona. We notice several ads repeated across iterations in Religion & Spirituality and Pets & Animals personas that do not seem to have any apparent personalization. For example, Amazon Eero WiFi (Figure 5b), Amazon Kindle, and Swarovski ads exclusively appeared

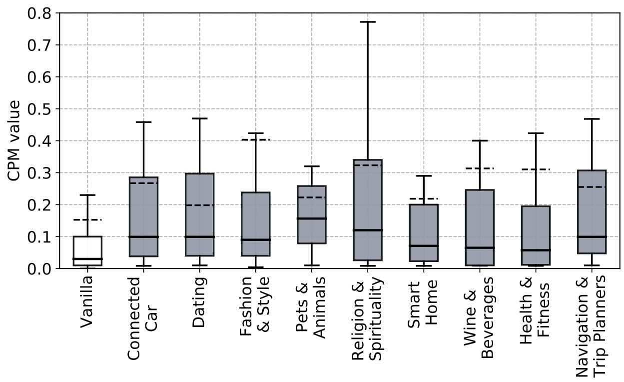
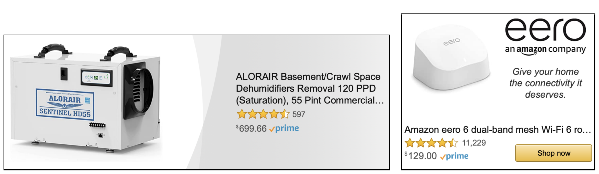


Figure 4: CPM values across vanilla (control) and interest (treatment) personas on common ad slots with user interaction. Solid and dotted lines in bars represent median and mean, respectively. Vanilla persona does not involve user interaction.

Persona	p-value	Adjusted p-value	Effect size
Connected Car	0.003	0.026	0.354
Dating	0.006	0.030	0.363
Fashion & Style	0.010	0.040	0.319
Pets & Animals	0.005	0.030	0.428
Religion & Spirituality	0.004	0.030	0.356
Smart Home	0.075	0.225	0.210
Wine & Beverages	0.083	0.225	0.192
Health & Fitness	0.149	0.225	0.139
Navigation & Trip Planners	0.002	0.015	0.410

Table 9: Statistical significance between vanilla (control) and interest (treatment) personas. p-value is computed through Mann-Whitney U test and adjusted through Holm-Bonferroni method. Effect size is rank-biserial coefficient.



(a) Dehumidifier ad in Health & Fitness

(b) Eero WiFi ad in Religion & Spirituality

Figure 5: Unique and repeated ads in interest personas.

on 12, 14, 2 times across 8, 4, and 2 iterations, respectively in Religion & Spirituality persona. Similarly, PC files copying/switching software ad appeared 4 times in 2 iterations in Pets & Animals persona.

Persona	Advertised products
Health & Fitness	Dehumidifier, Essential oils
Smart Home	Vacuum cleaner, Vac. clean. accessories
Religion & Spirituality	Wifi router, Kindle, Swarovski
Pets & Animals	PC files copying/switching software

Table 10: Ads from Amazon on interest personas. Green represents unique ads with apparent relevance to the persona. Yellow represents unique ads that repeat across iterations but do not have any apparent relevance to the persona.

6.6 Ad image analysis for personas with only skill installation

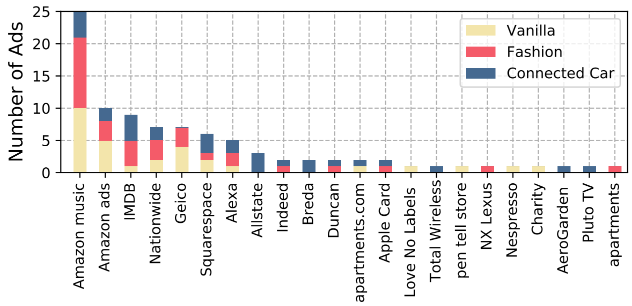
We manually analyze a total of 35 and 117 ads from installed skill vendors in their respective personas and Amazon with only skill installation. Out of the 35 ads from installed skills vendors, 16, 9, 2, 2, 1, and 1 are from Microsoft, SimpliSafe, Ring, and SharkNinja in Smart Home persona, respectively. Each of the Samsung, LG, and ATT also have a single ad in in Smart Home persona. 1 Honda and Dodge ad appears in Connected Car and 1 Starbucks ad appears in Wine & Beverages. Similar to personas with user interaction, none of the ads from installed skills vendors are exclusive to the personas where their skills are installed, which indicates that these ads do not reveal obvious personalization.

In contrast to ads from Amazon targeted to personas with user interaction, none of the 117 Amazon ads targeted to personas with only skill installation seem to be personalized. Only two ads are unique to Health & Fitness persona, i.e., an ad for an electric toothbrush appearing once and an ad for an air fryer toaster appearing 4 times but lack an apparent relevance to the persona, as per our rubric, i.e., Health & Fitness persona does not have any skills related to electric tooth brush or an air fryer toaster oven.

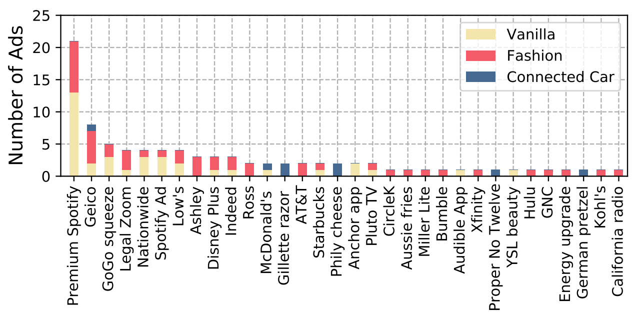
6.7 Audio ads are likely personalized

Next, we take a preliminary look at the audio ads directly collected on Amazon Echos from Amazon Music, Spotify, and Pandora. Table 11 shows the distribution of ads on each audio-streaming skill for tested personas. Since we record the same length of audio for each skill on each persona, we surmise that differences in the number of ads streamed across personas on the same skill, signal differences in advertiser interest [60]. For instance, as shown in Table 11, the number of ads on Spotify for the Connected Car persona is a fifth of the number of ads for the other two personas. We speculate that this considerable difference stems from the lower interest of advertisers to stream ads for this persona.

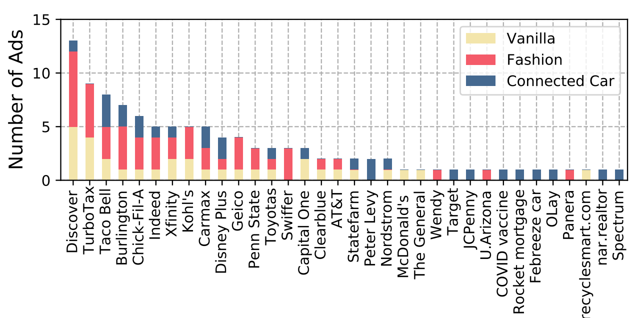
We also manually label the products advertised in order to look for evidence of obvious personalization (as we do in Section 6.5 for web ads). In this case, we only consider



(a) Audio ads on Amazon Music



(b) Audio ads on Spotify



(c) Audio ads on Pandora

Figure 6: Distribution of audio ads across Amazon Music, Spotify, and Pandora.

audio ads streamed twice or more, as repetitions may signal a stronger interest by the advertiser.

Figure 6 present the distribution of ads across Amazon Music, Spotify and Pandora. We find potential preliminary evidence of audio ad personalization for the Fashion & Style persona. Some advertising brands, such as Ashley and Ross on Spotify and Swiffer Wet Jet on Pandora, are exclusively streamed for Fashion & Style persona. Further, on Pandora, clothing brands such as Burlington and Kohl's appear much more frequently for the Fashion & Style persona than they do for other personas. We do not find similar patterns for the Connected Car persona, with the sole exception of Febreeze car on Pandora. We speculate that this persona does not reveal valuable information to audio ad vendors (unlike on the web,

Persona	Amazon	Spotify	Pandora
Connected Car	33.33% (31)	8.99% (8)	26.17% (28)
Fashion & Style	34.41% (32)	50.56% (45)	43.92% (47)
Vanilla	32.26% (30)	40.45% (36)	29.91% (32)
Total	100% (93)	100% (89)	100% (107)

Table 11: Fraction and count of ads on each audio-streaming skill per persona.

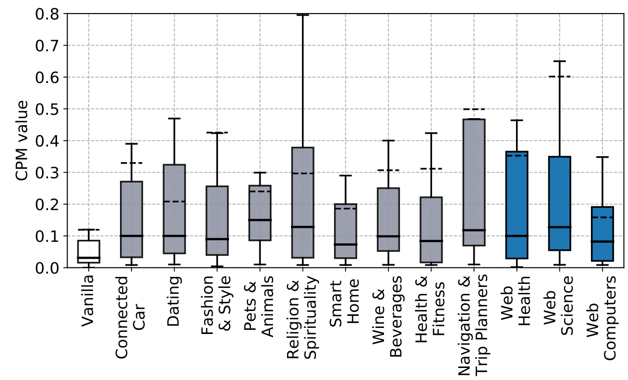


Figure 7: CPM values across vanilla, Echo interest, and web interest personas on common ad slots. Solid and dotted lines in bars represent median and mean, respectively.

see Section 6.5), as streaming music while driving a car is a widely popular activity. We also note that a large chunk of ads (16.61% of total ads) on Amazon Music and Spotify advertise the premium version of these two streaming services.

6.8 Echo interest personas are targeted similar to web interest personas

We compare Echo interest personas with web interest personas. Comparing Echo interest personas with web interest personas will allow us to draw parallels with the standard data usage and sharing on the web. Figure 7 presents the bidding values for Echo interest and web interest personas. It can be seen from the figure that there are no discernible differences between Echo interest and and web interest personas. We further conduct Mann-Whitney U test of statistical significance to validate our observation. Our null hypothesis is that the bid distributions of Echo interest personas are similar to web interest personas. We reject the null hypothesis if the p-value is less than 0.05. Table 12 shows the statistical significance between Echo interest and web personas. It can be seen from the table that for all persona combinations, except for Navigation & Trip Planners and web computers, there are no significant differences between Echo and web interest personas. We conclude that the voice data leaked through smart speakers and browsing data leaked through web, leads to similar amount of targeting.

Persona	p-value		
	Health	Science	Computers
Connected Car	0.857	0.752	0.243
Dating	0.910	0.722	0.162
Fashion & Style	0.964	0.586	0.277
Pets & Animals	0.600	0.691	0.059
Religion & Spirituality	0.815	0.976	0.125
Smart Home	0.504	0.147	0.879
Wine & Beverages	0.949	0.559	0.357
Health & Fitness	0.543	0.234	0.767
Navigation & Trip Planners	0.206	0.460	0.021

Table 12: Statistical significance between Echo interest (persona column) and web interest (Health, Science, Computers columns) personas. p-value is computed through Mann-Whitney U test.

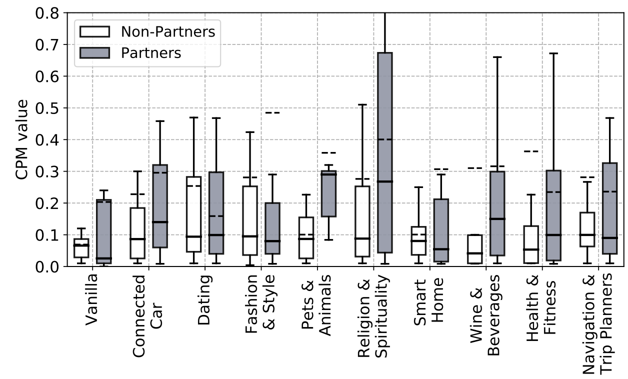


Figure 8: Bid values across personas on common ad slots distributed by Amazon's advertising partners.

Persona	Partner		Non-partner	
	Median	Mean	Median	Mean
Connected Car	0.140	0.296	0.086	0.228
Dating	0.099	0.159	0.094	0.254
Fashion & Style	0.080	0.485	0.095	0.281
Pets & Animals	0.290	0.358	0.087	0.101
Religion & Spirituality	0.268	0.400	0.088	0.276
Smart Home	0.054	0.307	0.080	0.101
Wine & Beverages	0.150	0.316	0.041	0.310
Health & Fitness	0.099	0.235	0.053	0.363
Navigation & Trip Plan.	0.090	0.236	0.100	0.281
Vanilla	0.025	0.203	0.352	0.066

Table 13: Median and mean bid values for personas from Amazon's partner and non-partner advertisers.

7 Data Sharing Analysis

In this section, we analyze the potential sharing of smart speaker interaction data from Amazon and third-party skills.

7.1 Some advertisers sync their cookies with Amazon and bid higher than non-cookie syncing advertisers

To target personalized ads, advertisers share user data with each other. Typically unique user identifiers, e.g., cookies, are shared at the client side with cookie syncing and user interest data is synced at the server side [58]. We analyze cookie syncing instances that involve Amazon advertising services in the network traffic captured while collecting ads (Section 3.3). We note that 41 third parties sync their cookies with Amazon across all Echo interest personas. Surprisingly, Amazon did not sync its cookies with any advertiser.[8] The one sided cookie-syncs could be explained by Amazon advertising's recent services for central identity resolution [88].

To infer potential data sharing by Amazon, we compare and contrast the bid values by Amazon's partners (i.e., cookie syncing advertisers) and non-partner advertisers. Figure 8 presents the bid values on common ad slots by Amazon's partner and non-partner advertisers. We note that both median and mean bid values from partners are high for 6 and 7 personas as compared to bids from non-partners, respectively. Median bid values are as much as 3× higher for Pets & Animals, Religion & Spirituality, and Wine & Beverages personas, while mean bid values are 3× higher for Pets & Animals, Smart Home, and vanilla personas. It is noteworthy that Amazon's advertising partners further sync their cookies with 247 other third parties, including advertising services. Such cookie syncs may lead to the propagation of user data in the advertising ecosystem.

[8]We analyze the OpenWPM datasets released by prior work [69] to validate that Amazon's cookie syncing behavior is not unique to our dataset.

7.2 It is unclear whether skills play a role in targeting of personalized ads

We try to quantify Amazon's and skills' role in higher bids and targeting of personalized ads. Since all interactions are mediated through Amazon, Amazon has the best vantage point to infer personas' interests and target personalized ads. Specifically, all voice inputs are interpreted by Amazon and most network requests are routed to/through Amazon (Table 2 and Figure 2). Amazon is also logged in to each persona and it can access its cookies to uniquely identify each persona. In fact, Section 5 and 6.5 already show that Amazon infers users' advertising interests from their voice data and uses the inferred interests to target personalized ads to users. We also note that Smart Home, Wine & Beverages, and Navigation & Trip Planners, personas do not contact any non-Amazon services but still receive high bid values, as compared to vanilla persona. Amazon also infers discernible interests for the Smart Home persona (Table 6). These results suggest that Amazon plays a crucial, if not a sole, role in higher bids and targeting of personalized ads.

In contrast, skills can only rely on persona's email address – if allowed permission, IP address – if skills contact non-Amazon web services, and Amazon's cookies – if Amazon collaborates with the skills, as unique identifiers to reach to personas. Though we allow skills to access email address, we do not log in to any online services (except for Amazon), thus skills cannot use email addresses to target personalized ads. Skills that contact non-Amazon web services and skills that collaborate with Amazon can still target ads to users. However, we note that only a handful (9) of skills contact few (12) advertising and tracking services (Table 2 and Figure 2), which cannot lead to mass targeting. Similarly, we note that none of the skills re-target ads to personas (Section 6.5), which implies that Amazon might not be engaging in data sharing partnerships with skills. Despite these observations, we still cannot rule out skills involvement in targeting of personalized ads.

8 Privacy Policy Analysis

In this section, we analyze the consistency between the data collection practices of Amazon and skill vendors with their privacy policies.

8.1 Collecting privacy policies

We obtain Amazon privacy policy from its website [12], which describes policies for most of Amazon's services, including Alexa. We download skills' privacy policies from *Developer Privacy Policy* link on the skill installation page. Recall from Section 3.1.1 that we experiment with 450 skills, i.e., top-50 skills from nine categories. Among the 450 skills, only 214 (47.6%) skills provide privacy policy links on their installation pages. The percentage is higher than the statistics reported by prior work [73], which identified that only 28.5% of the all skills provide a privacy policy link. We surmise that it could be because we investigate popular skills. Unfortunately, only 188 skills out of 214 provide a valid privacy policy link. Further, among the 188 obtained privacy policies, 129 do not even mention the word "Alexa" or "Amazon" in their text. We manually read many of the privacy policies, and notice that they are mostly generic and apply to several products from the same developer. Thus, they do not seem to be specific to Alexa skills.

8.2 Network traffic vs. privacy policies

We adapt and use PoliCheck [56] to check the consistency of data flows found in the network traffic with the statements declared in the corresponding privacy policy. PoliCheck has been previously applied to mobile app traffic [56], traffic from VR headsets [86], as well as to voice assistants [73]. However, in [73], data flows were extracted not from actual network traffic (as we do in this paper), but from the permissions of skills [73]. PoliCheck extracts ⟨*data type, entity*⟩ tuples from the data flows in the network traffic and from the disclosure text in statements in the privacy policies, and checks the consistency of the two. Based on the consistency between network traffic and policy tuples, PoliCheck assigns each data flow one of the following labels: (1) *clear*, when the data flow exactly matches a statement; (2) *vague*, when the data flow matches a statement in broader terms (e.g., the term "identifier" used for user ID); (3) *ambiguous*, when there are contradicting statements about a data flow; (4) *incorrect*, when the data flow corresponds to a statement that states otherwise, or (5) *omitted*, when there are contradicting statements about the data flow. For example, for *Sonos* skill [48] sending voice input to Amazon endpoints, PoliCheck will extract the ⟨*voice, amazon*⟩ network traffic tuple. From Sonos privacy policy, that mentions: *"The actual recording of your voice command is then sent to the voice partner you have authorized to receive such recording (for example, Amazon)."*, PoliCheck will extract the same ⟨*voice, amazon*⟩ policy tuple. Since the tuple from the network traffic matches the policy tuple, PoliCheck will assign the data flow a *clear* label.

Ideally, for each skill, we would run PoliCheck on the unencrypted network traffic collected from the AVS Echo to extract the skill's data flows and check them against the statements in the skill's privacy policy. However, due to the limitations of the AVS Echo, it only communicates with Amazon endpoints (see Section 3.2), we perform consistency analysis for each of the two pieces of information in the tuple. First, we adapt PoliCheck to perform the analysis only on the endpoints found in the encrypted traffic collected from the Amazon Echo. Second, we adapt PoliCheck to perform the analysis on the data types found in the unencrypted network traffic collected from the AVS Echo.

Endpoint analysis on encrypted traffic. Since we cannot extract data types from encrypted network traffic on commercial Amazon Echos, we modify PoliCheck to only validate the consistency of endpoint organizations contacted by skills and Amazon with their privacy policies. We update PoliCheck's entity ontology by inspecting the network traffic and including observed endpoints, which we then map to their organization using the methodology described in Section 3.2. Based on the service offered by the organization, it is assigned one or more categories from *platform provider*, *voice assistant service*, *analytic (tracking) provider*, *advertising network*, and *content provider*. These categories are derived from the original PoliCheck's entity ontology and terms found in the privacy policies. We visit the website of each organization to determine the service offered by it. *Platform provider* and *voice assistant service* labels are only assigned to Amazon. We also update consistency disclosure definitions. Specifically data flows are referred as (1) *clear*, when the endpoints are disclosed in the privacy policy using the exact organization name; (2) *vague*, when the endpoint is disclosed vaguely using category names or *third party*; and (3) *omitted*, when the endpoint is not disclosed at all. We do not use *ambiguous* and *incorrect* disclosures because a contradiction cannot be determined without considering data types. We label an endpoint as *no policy* when the skill does not provide a privacy policy.

Disclosure of platform-party collection. Table 14 presents the result of our endpoint analysis. It can be seen from the Table that only 10 privacy policies clearly indicate that personal information is collected by Amazon. For example, the *Sonos* skill [48] clearly states that voice recording is collected by Amazon. Furthermore, we find that 136 skills, vaguely disclose that their network traffic may go to Amazon. For example, the *Harmony* skill [29]) privacy policy mentions sending data but without referring the name of the entity: *"Circle products may send pseudonymous information to an analytics tool, including timestamps, transmission statistics, feature usage, performance metrics, errors, etc."*

Disclosure of first-party collection. We find that 32 skills connect to non platform-party endpoints. Among them, 10 provide privacy policies and only six have at least one clear or vague disclosure. The only two clearly disclosed first-party endpoints are in the privacy policies of the *YouVersion Bible* [53] and *Garmin* [27] skills: they correspond to the organizations that are the developers of the skills.

Disclosure of third-party collection. Many skills rely on third-party organizations, e.g., Liberated Syndication, Podtrac, Spotify and Triton Digital, which provide audio content distribution and tracking/advertising services. However, only a few skills disclose data collection and sharing with third-party organizations in their privacy policies, and when they do, they use vague terms. For example, the *Charles Stanley Radio* skill [18] uses the term *"external service providers"* to refer to third-party organizations in its privacy policy. Another example is the *VCA Animal Hospitals* skill that uses the blanket term *"third-parties"* to refer to all third-party organizations in its privacy policy [52].

Data types analysis on unencrypted traffic. We adapt PoliCheck to perform consistency analysis on the data types found in the unencrypted traffic collected from the AVS Echo. Following suit from prior work [73, 86], we rebuild PoliCheck's data ontology to include new data types, such as *voice recording*, that are unique to voice assistants and ignore vague terms, such as *"pii"*, *"user info"*, and *"technical info"*, as data types. Using the adapted PoliCheck, we check the consistency of data flows with disclosures. Similar to the endpoint analysis, the disclosure type of each data flow can be *clear*, *vague*, or *omitted* depending on whether the data type is disclosed using the exact term, vague term, or not disclosed at all.

Disclosure of data types in skill's privacy policies. Table 1 presents the result of our data types analysis using PoliCheck. It can be seen from the Table that 83 skills have at least one clear or vague disclosure. Among these skills, only 20 and 11 skills clearly disclose the collection of voice recordings and customer IDs. We note that 174 skills, despite providing privacy policies, do not disclose the collection of data types observable in their network traffic. For example, *Prayer Time* skill's privacy policy states the collection of browser version and visited pages, which only apply to their websites—indicating that the privacy policy is applied to several products as discussed in Section 8.1—and are not observable in the skill's network traffic at all.

Disclosure of data types in Amazon's privacy policy. As noted in Section 8.1, only 59 skills mention Amazon or Alexa in their privacy policies. Among these skills, 10 of them explicitly provide a link to Amazon's privacy policy or terms of use. Using PoliCheck, we find that including a link to Amazon's privacy policy makes all data flows to be either clearly or vaguely disclosed, depending on the terms used to disclose the data types in Amazon's privacy policy. This finding aligns with [86] that reports similar conclusions for the apps in the Oculus VR ecosystem. As all skills are built using Amazon's SDK and run on its platform, it is reasonable to assume that Amazon performs data collection as well. Unfortunately, most developers neither disclose this in their privacy policies, nor provide a link to Amazon's privacy policy.

Validation of PoliCheck results. To validate the correctness of PoliCheck when applied to skills, we manually inspect data flows from 100 skills that have a privacy policy, and check the consistency of these data flows with respect to the corresponding statements in the privacy policy. Following the methodology to validate PoliCheck results performed in [56, 73, 86], we consider *multi-class classification*. Similarly to [86], we assess the performance of the multi-class classification using micro- and macro-averaging. Thus, we obtain 87.41% micro-averaged precision, recall and F1-score. We also obtain the macro-averaged precision, recall, and F1-score as 93.96%, 77.85%, and 85.15% respectively.

9 User Perception Analysis

We have thus far shown that Amazon infers user interests from voice data and uses those interests to target ads to users. In this section, we conduct a user study to analyze users' expectations and comfort about the use of their data. Specifically, we try to answer two main questions: *(i)* based on Amazon's public statement (i.e., we "do not use voice recordings to target ads" [85]), do users expect that their voice recordings or information derived from their voice recordings would be used by Amazon for ad targeting, and *(ii)* whether users are comfortable with the use of interests inferred from their voice recordings or information derived from voice recordings for ad targeting. Our IRB-approved survey is included in Appendix 11.2. We conduct the user study on Prolific [41] with 393 participants in the US. 34.09% and 44.52% of our participants are between the age of 18–30 and 30–45, and 81.93% of the participants have non IT jobs (e.g., Software Engineer, Data Scientist). 51.90%, 45.03%, and 2.03% participants are male, female, and non-binary, respectively. Remaining participants did not share their gender. 60.95% of the participants own a smart speaker from Amazon, Google, or Apple.

Based on Amazon's statement, majority of the participants (63.10%) do not think that Amazon will use their voice recordings or the information derived from their voice recordings,

Endpoint Organization	Categories in the Ontology	Contacted Skills
Amazon Technologies, Inc.	analytic provider, advertising network, content provider, platform provider, voice assistant service	AAA Road Service , Salah Time , My Dog , My Cat , Outfit Check! , Pet Buddy , Rain Storm by Healing FM , Single Decade Short Rosary , Islamic Prayer Times , Sonos , 136 skills , 42 skills , 258 skills
Chartable Holding Inc	analytic provider, advertising network	Garmin , Makeup of the Day , My Tesla (Unofficial)
DataCamp Limited	content provider	Relaxing Sounds: Spa Music , Comfort My Dog , Calm My Cat
Dilli Labs LLC	content provider	VCA Animal Hospitals , EcoSmart Live , Dog Squeaky Toy , Relax My Pet , Dinosaur Sounds , Cat Sounds , Hush Puppy , Calm My Dog , Calm My Pet
Garmin International	content provider	Garmin
Liberated Syndication	analytic provider, advertising network	Calm My Pet , Al's Dog Training Tips
National Public Radio, Inc.	content provider	Makeup of the Day , Men's Finest Daily Fashion Tip
Philips International B.V.	content provider	Say a Prayer , Angry Girlfriend
Podtrac Inc	analytic provider, advertising network	Garmin , Gwynnie Bee , Genesis , Men's Finest Daily Fashion Tip , Love Trouble , Makeup of the Day , Dating & Relationship Tips
Spotify AB	analytic provider, advertising network	Gwynnie Bee , Genesis , Dating and Relationship Tips and advices , Makeup of the Day , Men's Finest Daily Fashion Tip , Love Trouble
Triton Digital, Inc.	analytic provider, advertising network	Garmin , Charles Stanley Radio
Voice Apps LLC	content provider	Prayer Time , Charles Stanley Radio , Morning Bible Inspiration , Holy Rosary , meal prayer , Halloween Sounds , Bible Trivia
Life Covenant Church, Inc.	content provider	YouVersion Bible , Lords Prayer

Table 14: Endpoint organizations observed in the network traffic from skills run on the Amazon Echo: only 32 skills exhibit non-Amazon endpoints. Skills highlighted in green use the exact organization name in the statement that discloses data collection and sharing by the endpoint. Skills highlighted in yellow use *third party* or other vague terms. Skills highlighted in red do not declare the contacted endpoint at all. Skills highlighted in gray do not provide a privacy policy.

to target ads. However, 29% of the participants still think that Amazon may use the information derived from voice recordings, 2.54% of participants think that Amazon may still use their voice recordings, and 5.34% of participants think that Amazon may still use their voice recordings or information derived from voice recordings, to targeted ads. It is important to note that these choices were listen as mutually exclusive options in the question. Users might believe that because of the general mistrust towards big tech, as expressed by one of the study participant: *Even when companies say they don't use your (specific) voice, they use ALL your data - including demographics and log in and time questions asked - so they might as well use your actual voice recording!*

We also find that participants can differentiate between voice recordings and the information derived from voice recordings (Questions 1 & 2 in Appendix 11.2). Participants feel slightly less uncomfortable when interests are inferred from the information derived from voice recordings (in contrast to directly from voice recordings) and subsequently used for ad targeting. Specifically, 72.25% of the participants would be extremely to somewhat uncomfortable with the ad targeting based on the interests inferred from their voice recordings. Whereas, 65.38% of the participants would be extremely to somewhat uncomfortable with the ad targeting based on the interests inferred from the information derived from their voice recordings. However, as we mention in Section 2, that to derive information from voice recordings, voice recordings need to be processed first. One participant precisely makes that observation: *The dilemma is people might be okay with information derived from their voice recordings being used in data-processing. But in order to process into suitable information the voice recording has to be synthesized and utilized by the company. So there's really no way around that privacy being breached. I think there'd be room for an enterprising company to come out with a smart-speaker that is designed first and foremost around privacy.*

10 Concluding Remarks

Takeaway. In this paper, we have audited the data collection, usage, and sharing practices in the Amazon smart speaker ecosystem. Our results indicate that *(i)* Amazon Echo user interactions are tracked by both Amazon and third-parties, *(ii)* Amazon uses Amazon Echo interactions data for ad targeting on-platform (e.g., audio ads) and off-platform (e.g., web ads), and *(iii)* Amazon computed user interests from voice data in a way that was inconsistent with their public statements. In many instances, Amazon and skills did not clearly disclose their data collection practices in their privacy policies. Furthermore, several skills did not provide any privacy policy or did not reference the platform's privacy policy. Given these findings, there is a clear need for increased transparency—by using auditing tools such as ours—on the practices of voice assistant platforms and third parties operating on them. The propagation of user data beyond the initial platform to the web is particularly alarming, as are the violations of privacy policies—which, as we show, are limited in scope, vague, and often even nonexistent for third parties.

Deployment. Our auditing framework and results may be useful to several stakeholders, including Amazon and skill developers (for internal privacy audits), policymakers (for crafting and effectively enforcing regulation), and users (as an incentive to guard their privacy using available tools). Upon publication we will release our code and data. We have shared our findings with the US Federal Trade Commission (FTC).

10.1 Possible Defenses

Improved transparency and control for users. Smart speakers users want to know what data is being collected, how that data is being used, and by whom. Our work suggests the need for greater transparency for users about the answer to these questions, as well as better control. Such transparency and control might come through a redesign of the platform itself (e.g., improved privacy-related UX, system-level enforcement with information flow control) or through external audits (such as with our framework) and external controls (either technical—e.g., network traffic filtering—and/or policy-based). For example, Amazon Echos are equipped with a debug interface [51]. Having such interface unlocked for developers and auditors would reveal the actual data being shared. Another example of a possible user defense is to selectively block network traffic that is not essential for the skill to work (e.g., using an approach similar to [74]).

Limiting voice interaction data. Even if the skills do not receive the actual voice recordings, smart-speaker platform does, since it has to transcribe them. Voice recordings do not only convey the command, but also other personal characteristics of the speakers (e.g., emotion, health, accent, etc. [84]). We can limit the sharing of this additional data by offloading the wake-word detection and transcription functions of the Alexa platform with offline tools such as [39, 43], and just send to the Alexa platform the transcribed commands using their textual API with no loss of functionality. Data sharing to only one vendor could also be limited by allowing users an option to install voice assistants from their preferred vendor, similar to mobile devices.

10.2 Parallels with Other IoT Platforms

Related platform-agnostic IoT works. Several IoT works have measured network traffic to detect data sharing. For example, [68, 74, 75, 81, 82] have shown that tracking is common in several IoT platforms, regardless of the presence of specific apps/skills. A difference between our findings and the ones of the above works is that Amazon smart speakers in our study contact additional endpoints from Amazon, skills vendors, and third-parties that have never been reported before. For example, with respect to the endpoints reported in a 2021 study [74], we have observed 4 new Amazon domains (`acsechocaptiveportal.com`, `amazon-dss.com`, `a2z.com`, `amazonalexa.com`.), 2 skills-specific endpoints (see *skills* row in Table 2) and 12 new third-party endpoints (see *third party* row in Table 2). A possible explanation could be the change in Amazon's ecosystem since it was last studied, e.g., `api.amazonalexa.com` may have replaced `api.amazon.com`, no longer contacted.

Related platform-specific IoT works. As compared to prior work on smart TVs [76, 87] and VR headsets [86], we have found less data tracking activity on smart speakers. However, on and off platform ad targeting indicates that data sharing still happens. A possible explanation could be the server-side data sharing from smart speaker platform for advertising purposes.

Generalization to other IoT platforms. Since indirect data sharing may happen in other IoT platforms as well, we envision that such platforms, including the ones already analyzed in prior work, may benefit from our approach for measuring data collection, usage, and sharing. For example, smart TV and VR platforms are amenable to our approach since we can collect network traffic, measure advertising and tracking, and check privacy policy compliance.

References

[1] Air Quality Report. `https://www.amazon.com/ICM-Air-Quality-Report/dp/B01EOFCHMA/`.

[2] Alexa and alexa device faqs. `https://www.amazon.com/gp/help/customer/display.html?nodeId=201602230`.

[3] Alexa blogs: Advertising and alexa. `https://developer.amazon.com/blogs/alexa/post/54c3a0f8-5b29-4071-acd7-2b832b860c83/advertising-and-alexa`.

[4] Alexa-hosted Skills. `https://developer.amazon.com/en-US/docs/alexa/hosted-skills/build-a-skill-end-to-end-using-an-alexa-hosted-skill.html`.

[5] Alexa privacy hub. `https://www.amazon.com/Alexa-Privacy-Hub/b?ie=UTF8&node=19149155011`.

[6] Alexa Skill Certification Requirements. https://developer.amazon.com/en-US/docs/alexa/custom-skills/certification-requirements-for-custom-skills.html.

[7] Alexa Skills Policy Testing. https://developer.amazon.com/en-US/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html.

[8] Alexa Skills Privacy Requirements. https://developer.amazon.com/en-US/docs/alexa/custom-skills/security-testing-for-an-alexa-skill.html#25-privacy-requirements.

[9] Alexa Top Sites by Category. https://https://www.alexa.com/topsites/category.

[10] Amazon Music. https://music.amazon.com/.

[11] Amazon: Request your data. https://www.amazon.com/gp/privacycentral/dsar/preview.html.

[12] Amazon.com Privacy Notice. https://www.amazon.com/gp/help/customer/display.html?nodeId=GX7NJQ4ZB8MHFRNJ.

[13] Audio Ads – Create audio advertising campaigns. https://advertising.amazon.com/en-ca/solutions/products/audio-ads.

[14] AVS Testing and Certification Process. https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/product-testing-overview.html.

[15] Best music streaming service for 2022 - cnet. https://www.cnet.com/tech/services-and-software/best-music-streaming-service/. (Accessed on 04/01/2022).

[16] Build with Amazon's Newest Devices & Services. https://developer.amazon.com/en-US/alexa/devices/alexa-built-in.

[17] Charles Stanley Radio. https://www.amazon.com/In-Touch-Ministries-Charles-Stanley/dp/B07FF2QGXW/ref=sr_1_101?dchild=1&qid=1602785535&s=alexa-skills&sr=1-101.

[18] Charles Stanley Radio. https://www.amazon.com/dp/B07FF2QGXW/.

[19] 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.

[20] Creating and Managing List Skills FAQ. https://developer.amazon.com/de-DE/docs/alexa/custom-skills/list-faq.html.

[21] Crunchbase. https://www.crunchbase.com/.

[22] Dating and Relationship Tips and advices. https://www.amazon.com/Aaron-Spelling-Dating-Relationship-advices/dp/B07YCKFCCF/ref=sr_1_28?dchild=1&qid=1602782676&s=alexa-skills&sr=1-28.

[23] Distinction between user and person in skill requests. https://developer.amazon.com/en-US/docs/alexa/custom-skills/add-personalization-to-your-skill.html#distinction-between-user-and-person-in-skill-requests.

[24] DuckDuckGo Tracker Radar list of entities. https://github.com/duckduckgo/tracker-radar/tree/main/entities.

[25] Dyson. https://www.amazon.com/Dyson-Limited/dp/B06WVN7SHC/.

[26] Essential Oil Benefits. https://www.amazon.com/ttm-Essential-Oil-Benefits/dp/B074CNX3G8/.

[27] Garmin. https://www.amazon.com/dp/B075TRB4V5/.

[28] Genesis. https://www.amazon.com/Genesis-Motors-USA/dp/B01JXP09PI/ref=lp_14284820011_1_6?s=digital-skills&ie=UTF8&qid=1602832937&sr=1-6.

[29] Harmony. https://www.amazon.com/dp/B01M4LDPX3/.

[30] Header Bidding. https://admanager.google.com/home/resources/feature-brief-open-bidding/.

[31] Header Bidding (HBIX) 2021 Tracker. https://www.kevel.co/hbix/.

[32] Makeup of the Day. https://www.amazon.com/Xeline-Development-Makeup-the-Day/dp/B072N6BNB1/ref=sr_1_232?dchild=1&qid=1602773008&s=alexa-skills&sr=1-232.

[33] Men's Finest Daily Fashion Tip. https://www.amazon.com/Mens-Finest-Daily-Fashion-Tip/dp/B07CB3ZN6N/ref=lp_14284840011_1_5?s=digital-skills&ie=UTF8&qid=1602772432&sr=1-5.

[34] Men's Finest Daily Fashion Tip. https://www.amazon.com/Mens-Finest-Daily-Fashion-Tip/dp/B07CB3ZN6N/ref=lp_14284840011_1_5?s=digital-skills&ie=UTF8&qid=1602772432&sr=1-5.

[35] Module 2: Design an engaging voice user interface. https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/tutorials-code-samples/build-an-engaging-alexa-skill/module-2.

[36] pandora. https://www.amazon.com/Pandora-Media/dp/B07JBQZCRB.

[37] Pi-hole Blocklist. https://firebog.net/.

[38] Policy testing. https://developer.amazon.com/en-US/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html#advertising.

[39] Porcupine Wake Word. https://picovoice.ai/platform/porcupine/.

[40] Prebid. https://prebid.org/.

[41] Prolific. https://www.prolific.co.

[42] Real-time Bidding. https://developers.google.com/authorized-buyers/rtb/start.

[43] Rhasspy Voice Assistant. https://rhasspy.readthedocs.io/.

[44] Security Policy for Device SDKs. https://github.com/alexa/avs-device-sdk/security/policy.

[45] Selenium. http://docs.seleniumhq.org/.

[46] Setting up a Bridged Wireless Access Point. https://www.raspberrypi.com/documentation/computers/configuration.html#setting-up-a-bridged-wireless-access-point.

[47] Smart speaker devices installed base in the united states from 2017 to 2020. https://www.statista.com/statistics/794480/us-amazon-echo-google-home-installed-base/.

[48] Sonos. https://www.amazon.com/dp/B072ML3N6K/.

[49] Spotify. https://www.amazon.com/Spotify/dp/B07FK56GVY.

[50] Streaming music report sheds light on battle between spotify, amazon, apple, and google - the verge. https://www.theverge.com/2022/1/20/22892939/music-streaming-services-market-share-q2-2021-spotify-apple-amazon-tencent-youtube. (Accessed on 04/01/2022).

[51] Uncovering the Echo Dot's Hidden USB Port. https://hackaday.com/2019/08/15/uncovering-the-echo-dots-hidden-usb-port/.

[52] VCA Animal Hospital. https://www.amazon.com/dp/B07KYS1Y1X/.

[53] YouVersion Bible. https://www.amazon.com/dp/B017RXFNKY/.

[54] AMAZON. Amazon echo & alexa devices. https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=9818047011.

[55] ANALYTICS, S. S. Number of households with smart home products and services in use worldwide from 2015 to 2025. https://www.statista.com/statistics/1252975/smart-home-households-worldwide/.

[56] ANDOW, B., MAHMUD, S. Y., WHITAKER, J., ENCK, W., REAVES, B., SINGH, K., AND EGELMAN, S. Actions speak louder than words:{Entity-Sensitive} privacy policy and data flow analysis with {PoliCheck}. In *29th USENIX Security Symposium (USENIX Security 20)* (2020), pp. 985–1002.

[57] BARCELÓ-ARMADA, R., CASTELL-UROZ, I., AND BARLET-ROS, P. Amazon alexa traffic traces. *Computer Networks 205* (2022), 108782.

[58] BASHIR, M. A., ARSHAD, S., ROBERTSON, W., AND WILSON, C. Tracing information flows between ad exchanges using retargeted ads. In *25th USENIX Security Symposium* (2016).

[59] CHENG, L., WILSON, C., LIAO, S., YOUNG, J., DONG, D., AND HU, H. 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* (2020).

[60] COOK, J., NITHYANAND, R., AND SHAFIQ, Z. Inferring tracker-advertiser relationships in the online advertising ecosystem using header bidding. *arXiv preprint arXiv:1907.07275* (2019).

[61] COOK, J., NITHYANAND, R., AND SHAFIQ, Z. Inferring tracker-advertiser relationships in the online advertising ecosystem using header bidding. In *Privacy Enhancing Technologies Symposium (PETS)* (2020).

[62] DUBOIS, D. J., KOLCUN, R., MANDALARI, A. M., PARACHA, M. T., CHOFFNES, D., AND HADDADI, H. When speakers are all ears: Characterizing misactivations of iot smart speakers. *Proceedings on Privacy Enhancing Technologies* (2020).

[63] EDU, J., ARAN, X. F., SUCH, J., AND SUAREZ-TANGIL, G. Skillvet: Automated traceability analysis of amazon alexa skills.

[64] Englehardt, S., and Narayanan, A. Online Tracking: A 1-million-site Measurement and Analysis. In *ACM Conference on Computer and Communications Security (CCS)* (2016).

[65] Fowler, G. A. Alexa has been eavesdropping on you this whole time. https://www.washingtonpost.com/technology/2019/05/06/alexa-has-been-eavesdropping-you-this-whole-time/, 2019.

[66] Gary Horcher, KIRO 7 News. Woman says her amazon device recorded private conversation, sent it out to random contact. https://www.kiro7.com/news/local/woman-says-her-amazon-device-recorded-private-conversation-sent-it-out-to-random-contact/755507974/.

[67] Google. RTB - Cookie Matching. https://developers.google.com/authorized-buyers/rtb/cookie-guide.

[68] Huang, D. Y., Apthorpe, N., Li, F., Acar, G., and Feamster, N. IoT inspector: Crowdsourcing labeled network traffic from smart home devices at scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4*, 2 (2020), 1–21.

[69] Iqbal, U., Wolfe, C., Nguyen, C., Englehardt, S., and Shafiq, Z. Khaleesi: Breaker of advertising and tracking request chains. In *USENIX Security Symposium (USENIX)* (2022).

[70] iRobot. irobot home. https://www.amazon.com/iRobot-Home/dp/B06Y3PSHQ3.

[71] Jin, H., and Wang, S. Voice-based determination of physical and emotional characteristics of users, Oct. 2018. US Patent 10,096,319.

[72] Le Pochat, V., Van Goethem, T., Tajalizadehkhoob, S., Korczyński, M., and Joosen, W. Tranco: A research-oriented top sites ranking hardened against manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium* (2019), Internet Society.

[73] Lentzsch, C., Shah, S. J., Andow, B., Degeling, M., Das, A., and Enck, W. 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).

[74] Mandalari, A. M., Dubois, D. J., Kolcun, R., Paracha, M. T., Haddadi, H., and Choffnes, D. Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)* (2021).

[75] Mazhar, M. H., and Shafiq, Z. Characterizing smart home IoT traffic in the wild. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)* (2020), IEEE, pp. 203–215.

[76] Mohajeri Moghaddam, H., Acar, G., Burgess, B., Mathur, A., Huang, D. Y., Feamster, N., Felten, E. W., Mittal, P., and Narayanan, A. Watching you watch: The tracking ecosystem of over-the-top tv streaming devices. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (2019), pp. 131–147.

[77] NBC. Are Smart Speakers Planting Ads On Our Social Media Profiles? https://www.nbcmiami.com/news/local/are-smart-speakers-planting-ads-on-our-social-media-profiles/157153/.

[78] Olejnik, L., Tran, M.-D., and Castelluccia, C. Selling off privacy at auction. In *Network and Distributed System Security Symposium (NDSS)* (2014).

[79] Papadopoulos, P., Kourtellis, N., Rodriguez, P., and Laoutaris, N. If you are not paying for it, you are the product: How much do advertisers pay to reach you? In *Proceedings of the 2017 Internet Measurement Conference* (2017).

[80] Pro, A. P. Adobe premiere pro, 2022.

[81] Ren, J., Dubois, D. J., Choffnes, D., Mandalari, A. M., Kolcun, R., and Haddadi, H. Information Exposure for Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach. In *Proc. of the Internet Measurement Conference (IMC)* (2019).

[82] Saidi, S. J., Mandalari, A. M., Kolcun, R., Haddadi, H., Dubois, D. J., Choffnes, D., Smaragdakis, G., and Feldmann, A. A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. In *Proc. of the Internet Measurement Conference (IMC)* (2020).

[83] Shaban, H. Amazon alexa user receives 1,700 audio recordings of a stranger through 'human error'. https://www.washingtonpost.com/technology/2018/12/20/amazon-alexa-user-receives-audio-recordings-stranger-through-human-error/, 2018.

[84] Singh, R. *Profiling humans from their voice*. Springer, 2019.

[85] Times, N. Y. Hey, Alexa, What Can You Hear? And What Will You Do With It? https://www.nytimes.com/2018/03/31/business/

media/amazon-google-privacy-digital-assistants.html.

[86] TRIMANANDA, R., LE, H., CUI, H., TRAN HO, J., SHUBA, A., AND MARKOPOULOU, A. Ovrseen: Auditing network traffic and privacy policies in oculus vr. In *31st USENIX security symposium (USENIX security 22)* (2022).

[87] VARMARKEN, J., LE, H., SHUBA, A., MARKOPOULOU, A., AND SHAFIQ, Z. The TV is Smart and Full of Trackers: Measuring Smart TV Advertising and Tracking. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)* (2020).

[88] WILLENS, M. Amid post-cookie confusion, amazon plans to launch an identifier of its own. https://digiday.com/marketing/amid-post-cookie-confusion-amazon-explores-launching-an-identifier-of-its-own/amp/, 2021.

[89] YOUNG, J., LIAO, S., CHENG, L., HU, H., AND DENG, H. SkillDetective: Automated Policy-Violation detection of voice assistant applications in the wild. In *USENIX Security Symposium* (2022).

[90] ZHANG, J., PSOUNIS, K., HAROON, M., AND SHAFIQ, Z. Harpo: Learning to subvert online behavioral advertising. In *Network and Distributed System Security Symposium (NDSS)* (2022).

11 Appendix

11.1 Amazon's reaction to our research

Amazon's statement about our research and our response to Amazon's statement are published in news articles. We have redacted the details in the text below to avoid violating the double blind policy by leaking our identity.

11.1.1 Amazon's statement

Amazon gave the following response to a journalist query about our research:

> *Many of the conclusions in this research are based on inaccurate inferences or speculation by the authors, and do not accurately reflect how Alexa works. We are not in the business of selling data and we do not share Alexa requests with advertising networks. Similar to what you'd experience if you made a purchase on Amazon.com or requested a song through Amazon Music, if you ask Alexa to order paper towels or to play a song on Amazon Music, the record of that purchase or song play may inform relevant ads shown on Amazon or other sites where Amazon places ads. Customers can opt out of interest-based ads from Amazon at anytime on our website.*

Another journalist asked Amazon for a clarification about inaccuracies in our findings. Amazon's exchange with the journalist:

> *An Amazon spokesperson is telling me the company does not sell customers' personal information or share Alexa requests with advertising networks. The spokesperson is saying the report suggests that they do, hence why they're calling it "inaccurate".*

11.1.2 Our response

We welcome critiques of our research methodology and our findings, but Amazon's statement does not directly address our findings. Specifically, we find that Echo devices running Alexa skills communicate with advertising services (Section 4.2). We find that Amazon infers users' advertising interests from their Echo interactions (Section 6). We find that Amazon's advertising partners sync (share) cookies with Amazon, and that Amazon's partner advertisers bid more than non-partner advertisers to place ads for Echo personas (users) that install and interact with Alexa skills (Section 7.1). We also find that Amazon's and Echo skills' operational practices are often not clearly disclosed in their privacy policies (Section 8). We do not claim that Amazon sells customers' personal information neither do we claim that Amazon shares Alexa requests with advertising networks.

Most importantly, Amazon's statement tells users that it serves ads based on Echo interactions and that users can opt out of interest-based ads. This confirms our conclusion that it indeed uses interests inferred from users' interactions with Echos for behavioral advertising. Amazon does not refute our claims that it also shares users' interests with its advertising partners.

11.2 User Perception Analysis Questionnaire

We report the results of our user perception analysis in Section 9. The following is the list of questions from our IRB-approved survey questionnaire.

An overview of Amazon Echo smart speaker. Amazon's smart speakers are called Echo and they are powered by the Alexa voice assistant. Alexa is an artificial intelligence (AI) based voice assistant that responds to questions or commands that people ask of it orally. For example, you might place an Amazon Echo in your living room and speak to it to check the weather, play music, or control smart devices like lights.

When a user interacts with Amazon Echo, data about the interaction might be collected, stored, and used. That data might include the **voice recording** and/or **the information derived from voice recording**.

For a command issued to check air quality on Amazon Echo, following could be the voice recording and information derived from voice recording:
Voice recording: Alexa, ask the Air Quality app to give me the air quality in Seattle?
Information derived from voice recording: The user interacted with the Air Quality app to check weather in Seattle.

1. Consider a scenario where a user asks Amazon Echo for pain reducing remedies for diabetic patients. Which of the following would be the **voice recording**?
    - Alexa, ask My Diabetes Lifestyle app about reducing pain?
    - The user interacted with the My Diabetes Lifestyle app about pain reducing remedies/treatments.
2. Consider a scenario where a user asks Amazon Echo for pain reducing remedies for diabetic patients. Which of the following would be the **information derived from voice recording**?
    - Alexa, ask My Diabetes Lifestyle app about reducing pain?
    - The user interacted with the My Diabetes Lifestyle app about pain reducing remedies/treatments.
3. In a statement to the New York Times, Amazon said that: the company took "privacy seriously" and did "not use customers' voice recordings for targeted advertising."

    Assuming that Amazon is honest about this statement and abides by this policy, which of the following might happen when you interact with Amazon Echo:
    - Amazon may use your **voice recordings** to target ads.
    - Amazon may use **information derived from your voice recordings** to target ads.
    - Amazon will not use **voice recordings**, or **information derived from voice recordings**, to target ads.
    - Amazon may use **voice recordings**, and **information derived from voice recordings**, to target ads.
4. People may be comfortable or uncomfortable with the ads targeted based on the interests inferred from their **voice recordings** on smart speakers.

    On a scale of extremely uncomfortable to extremely comfortable, how comfortable would you be with Amazon targeting you personalized ads on websites (e.g., news and health related websites), based on the interests inferred from your **voice recordings** (e.g., interested in health related information)?
    - Extremely uncomfortable
    - Somewhat uncomfortable
    - Neither comfortable nor uncomfortable
    - Somewhat comfortable
    - Extremely comfortable
5. People may be comfortable or uncomfortable with the ads targeted based on the interests inferred from the **information derived from voice recording** on smart speakers.

    On a scale of extremely uncomfortable to extremely comfortable, how comfortable would you be with Amazon targeting you personalized ads on news and health related websites (e.g., news and health related websites), based on the interests inferred from the **information derived from voice recording** (e.g., interested in health related information)?
    - Extremely uncomfortable
    - Somewhat uncomfortable
    - Neither comfortable nor uncomfortable
    - Somewhat comfortable
    - Extremely comfortable
6. Some people read privacy policies while others do not. How often do you read privacy policies?
    - I often read them in detail.
    - I have one or twice read them in detail.
    - I glance through them.
    - I never really look at them.
7. Do you use a smart speaker in your home? Please select all that you use.
    - I do not use a smart speaker.
    - Amazon smart speaker (Amazon Echo, etc.)
    - Google smart speaker (Google Home, etc.)
    - Apple smart speaker (Apple HomePod, etc.)
    - Other
8. Any thoughts that you would like to share with us?