# 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 voice data to infer user interests and uses it 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 [52], such as Amazon Echo [51], 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 [59]. 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 [82]. 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 [56], [87], record users' private conversations without their knowledge [62], [63], and share users' conversations with strangers [81]. 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 [69].

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 voice 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, can allow 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 [43] and 200K third-party applications [60]). To address the first challenge, we set up a custom Raspberry RPi (RPi) router [42] to capture the endpoints contacted by Amazon Echo and as well as emulating an Amazon Echo by instrumenting Alexa Voice Service (AVS) SDK [15] and running it on a RPi (we call it AVS Echo) to capture collected data. Since our custom RPi router is unable to decrypt TLS traffic from unmodified Amazon Echos, we configure our AVS Echo to capture unencrypted network traffic.

To address the second challenge, we conduct controlled experiments where we intentionally expose 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. For example, 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 [77], [58], [55]. 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 with Amazon. These advertisers further sync their cookies with 247 other third parties, including advertising services.
2) *Is voice data used by either Amazon or third-party apps beyond purely functional purposes, such as for targeted advertising?* Amazon processes voice data to infer user interests. Our measurements indicate the usage of voice data 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' voice interactions seems to be inconsistent with their public statements [83], [75]. Similarly, more than 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 all of 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 [43]. Amazon's smart speakers are called Echo and they are powered by the Alexa voice assistant. 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 [60].

## 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], [6], and requires user permission to allow access to personal information, e.g., email, phone, location [18]. 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 [5]. However, prior research has shown that policy-violating skills can get certified [56] and thousands of skills on the Alexa marketplace violate platform policies [87].

Smart speakers also handle the particularly sensitive data that consists of users' voice input. The content of users' speech can reveal sensitive information (e.g., private conversations) and the voice signals can be processed to infer potentially sensitive information about the user (e.g., age, gender, health [82]). Amazon aims to limit some of these privacy issues through its platform design choices [4]. Specifically, to avoid snooping on sensitive conversations, voice input is only recorded when a user utters the *wake word*, e.g., Alexa. Further, only processed transcriptions of voice input (not the audio data) is shared with third party skills, instead of the raw audio [32]. However, despite these design choices, prior research has also shown that smart speakers often *misactivate* and unintentionally record conversations [59]. In fact, there have been several real-world instances where smart speakers recorded user conversations, without users ever uttering the wake word [63].

Smart speakers typically send voice input to cloud servers for processing (e.g., transcription), after which the data can be stored and shared with other parties. This raises two privacy concerns. First, since the potentially sensitive data from voice interactions is available to smart speaker vendors, they can use this data for targeting ads (as proposed in a recent Amazon patent [69]). Second, this data may be shared with other parties. For example, when a user interacts with a third party skill, the (processed transcriptions of) voice input is shared with the third party. In these cases, neither users nor Amazon have any visibility or control on the processing, sharing, and selling of users' interpreted voice input. Third party skills often do not publish their privacy policies, nor adhere to them even when they do [60].

### 2.3. Proposed Auditing Framework

To the best of our knowledge, prior work lacks an in-depth analysis of the collection, sharing, and usage of data in the Alexa smart speaker ecosystem. To fill this gap, we systematically analyze the data collection, sharing, and usage practices of Amazon's smart speaker platform including third party skills. We conduct controlled experiments where we intentionally expose user interests according to several personas, then observe the platform's subsequent behavior from three perspectives: *(i) network traffic* exchanged by smart speakers, *(ii) advertisements* served to personas, 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 observe communication between Amazon and some third parties, we otherwise find that the Amazon ecosystem uses an opaque communication model where encryption and proxying hide substantial amounts of information flows among devices, Amazon servers, and third parties.

*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 [27], [38]. 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 Alexa platform 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 [2] and typically requires rather high minimum ad spend commitment from advertisers [12].

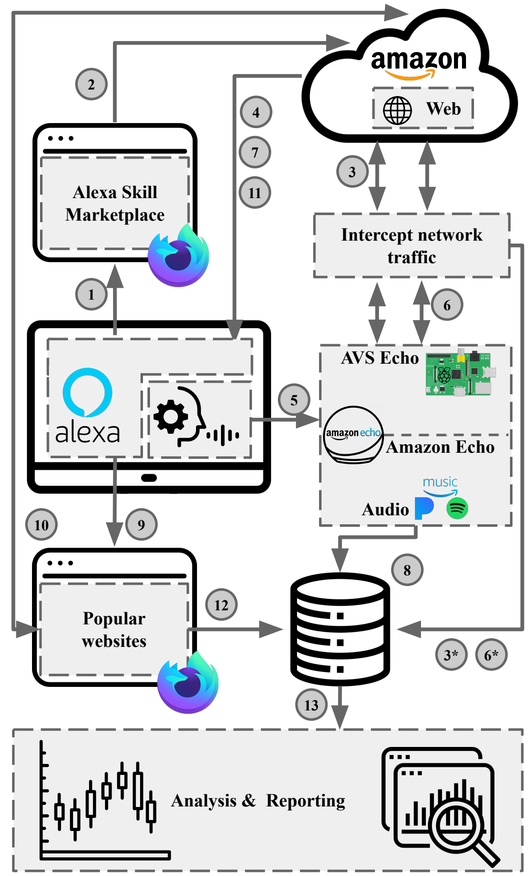



Figure 1: Approach overview: (1–4) we link Amazon Echo to the Alexa web companion app and visit Alexa skill marketplace to install skills, (5–8) we then interact with the installed skills by uttering sample invocation utterances listed in skill description, (9–11) we then visit popular websites while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

## 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 personas by installing and interacting with skills from nine different 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.

*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. We then use a Selenium [41] 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 released in [60] to extract top skills. 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 often also linking a physical IoT device, e.g., iRobot skill requires to link a robot vacuum cleaner with the skill [68].

*Skill interaction.* After installing each skill, we interact with it by programmatically uttering sample invocations. 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 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 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 [8], and are referred to as *web health*, *web science*, and *web computer* personas. We use OpenWPM [61], an open-source web measurement tool to prime web personas. Similar to interest personas, to avoid contamination across control personas, we configure each control persona 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. 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 [77], [76], [58].

## 3.2. Capturing network traffic

We capture outgoing and incoming network traffic, to and from, Amazon Echos to measure data tracking by Amazon and skills. 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 [42] 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, 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 tracked data. To enhance our coverage, we simulate an Echo device by instrumenting Alexa Voice Service (AVS) SDK [15] 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 [40]. Further, unlike commercial Amazon Echos that can communicate with Amazon and third-party endpoints, AVS Echo only communicates with Amazon.

*Inferring origin.* Both encrypted and unencrypted 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 [21], Crunchbase [19], 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 [76], [77], [58] has shown that the advertisers bidding behavior is influenced by their pre-existing 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 [27] allows to observe bid values at the client side, we collect ad bids and ad images on header bidding supported websites. To this end, we first identify top websites that support `prebid.js` [36], the most widely used implementation of header biding protocol [28], and then visit those websites to capture bids and ad images. We extend OpenWPM [61] to identify and capture data on `prebid.js` supported websites. To identify `prebid.js` supported websites, we crawl Tranco top websites list [70] 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 [76], [77]. We try to minimize the variability by keeping conditions consistent across personas. Specifically, we use identical hardware/software, collect bids at the same time, 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 [77] and advertiser may not bid on all ad slots across all personas. We also relatively compare bid values across personas because their 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 (6 times before interacting with skills and 25 times after interacting with skills) 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 [64], between Amazon and its advertising partners. Note that the network traffic captured while crawling is different from network 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 through audio ads, despite their shortcomings (mentioned in Section 2). We capture audio ads played on three audio-streaming skills: Amazon Music [9], Spotify [45], and Pandora [33]. We include Amazon Music to determine if Amazon (the platform operator) personalizes audio ads, while the other two are popular streaming services [46], [14] with over 10,000 reviews on the Alexa platform [45], [33]. Since ads are played at variable intervals in-between songs, we stream music for several hours. Specifically, we stream and record six hours of *top hit music* for each skill. We then automatically transcribe the recorded audio files [78] 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 5.4), 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 our methodology to collect and process 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—as with web ads, we 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

### 4.1. Amazon has the best vantage point to track user activity

Table 1 presents the list of domains contacted by skills. 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 (4 (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 [32]. Another possible explanation for a large number of network traffic flows to Amazon could be the hosting of skills on Amazon's platform [3]. *Garmin* [24]

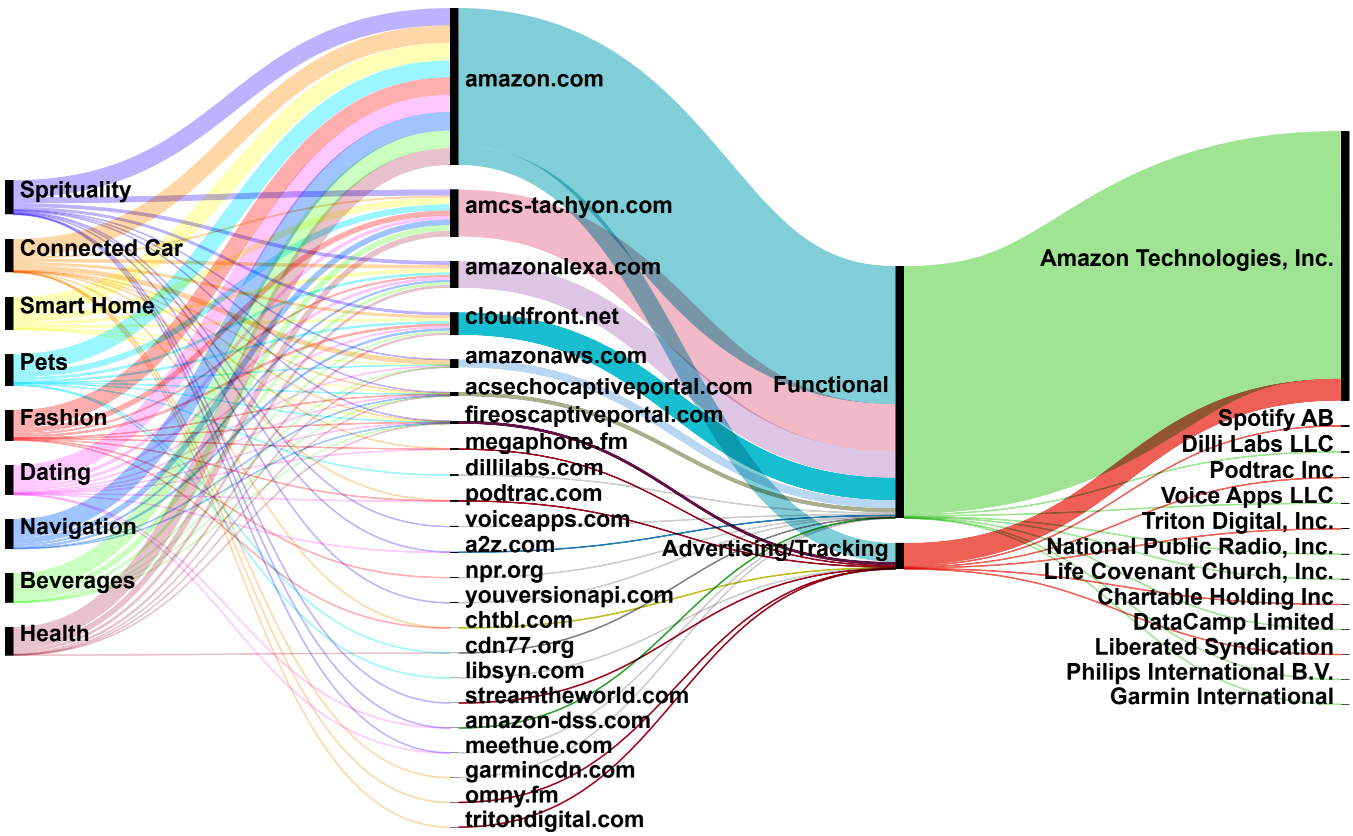



Figure 2: Network traffic distribution by persona, domain name, purpose, and organization

and *YouVersion Bible* [50] 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 1, 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 13 presents the details of data shared by skills. 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, 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

Several domains contacted by skills offer audio advertising and tracking services (rows highlighted in gray in Table 1). We rely on filter lists [34] and manual investigations to detect advertising and tracking services. Table 2 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 [54], 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* [25] and *Men's Finest Daily Fashion Tip* [31] with 398 and 13 reviews, contact these third-party advertising and tracking services. Six of such skills do not stream music, radio, podcast, or provide a flash briefing, which potentially violates Amazon's Alexa advertising policy that restricts non-streaming skills from playing ads [2]. 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 why these skills were not flagged during skill certification [13].

Table 3 and 4 further provide the distribution of advertising and tracking domains by personas and skills. From Table 3, 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 4, we note that skills contact several advertising and tracking services. The skill *Garmin* [24] even contacts as much as 4 advertising and tracking services.

***Takeaway.*** Amazon is in the best position to track user

| Org. | Domains | Skills |
|---|---|---|
| Amazon | *(11).amazon.com | 895 |
| | prod.amcs-tachyon.com | 305 |
| | api.amazonalexa.com | 173 |
| | *(7).cloudfront.net | 144 |
| | 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 |
| Skills | *(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 1: 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.

| Organization | Functional | Advertising & Tracking | Total |
|---|---|---|---|
| **Amazon** | 88.93% | 7.91% | 96.84% |
| **Skill vendor** | 0.17% | 0% | 0.17% |
| **Third party** | 1.49% | 1.50% | 2.99% |
| **Total** | 90.59% | 9.4% | 100% |

TABLE 2: Distribution of advertising / tracking and functional network traffic by organization.

activities because most traffic is mediated through Amazon. Even if users intend, they cannot interact with the skills without Amazon's involvement. We also note that six non-streaming skills send data directly from the smart speaker to advertising and tracking services, which could be a potential violation of Amazon's Alexa advertising policy [2].

| Persona | Advertising & Tracking | Functional |
|---|---|---|
| **Fashion & Style** | 9 | 4 |
| **Connected Car** | 7 | 0 |
| **Pets & Animals** | 3 | 11 |
| **Religion & Sprituality** | 3 | 8 |
| **Dating** | 5 | 1 |
| **Health & Fitness** | 0 | 1 |

TABLE 3: Count of advertising/tracking and functional third-party domains contacted by personas.

| Skill name | Advertising & Tracking |
|---|---|
| Garmin [24] | chtbl.com<br>traffic.omny.fm<br>dts.podtrac.com<br>turnernetworksales.mc.tritondigital.com |
| Makeup of the Day [29] | *(2).megaphone.fm<br>play.podtrac.com<br>chtbl.com |
| Men's Finest Daily Fashion Tip [30] | play.podtrac.com<br>*(2).megaphone.fm |
| Dating and Relationship Tips and advices [20] | play.podtrac.com<br>*(2).megaphone.fm |
| Charles Stanley Radio [16] | *(2).streamtheworld.com |

TABLE 4: 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.

| 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 |

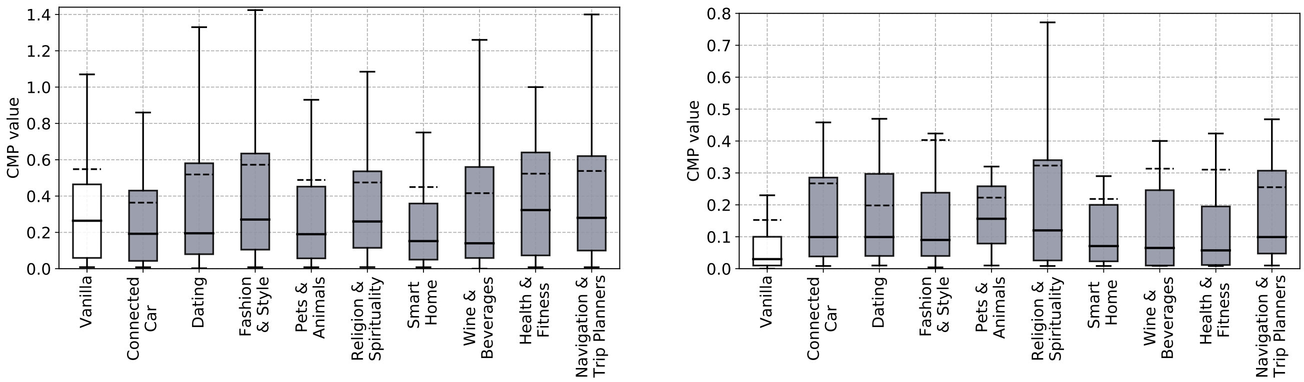
TABLE 5: Median and mean bid values (CPM) for interest (treatment) and vanilla (control) personas.

## 5. Ad Targeting analysis

### 5.1. User interaction leads to higher bid values

Figure 3 presents bid (CPM)[2] values across vanilla and interest personas on common ad slots without and with (Figure 3b) user interaction. It can be seen from Figure 3a that without user interaction, there is no discernible difference between vanilla and interest personas. Whereas, with user interaction, i.e., Figure 3b, the bid values are significantly higher for interest personas as compared to vanilla persona. Table 5 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.

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




(a) Bidding behavior without user interaction

(b) Bidding behavior with user interaction

Figure 3: CPM values across vanilla (control) and interest (treatment) personas on common ad slots without and with user interaction. Solid and dotted lines in bars represent median and mean, respectively.

| 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 6: Mean bid values without and with interaction across interest and vanilla personas that were collected close to each other.

*High bid values without user interaction.* The high bid values without user interaction could be explained by data collection during the holiday season, i.e., before Christmas 2021. To rule out the impact of holiday season, we compare the bids values without and with 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 holiday season. Table 6 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 discernable differences exist for without interaction configurations.

## 5.2. Interest personas have statistically higher bids than vanilla persona

We perform the Mann-Whitney U test to analyze whether interest personas receive significantly higher bids than vanilla persona. Our null hypothesis is that the bid distributions for interest personas are similar to vanilla persona. Whereas 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 $\geq$ 0.43 are considered small, medium, and large, respectively.

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

TABLE 7: Statistical significance between vanilla (control) and interest (treatment) personas. $p$-value is computed through Mann-Whitney U test. Effect size is rank-biserial coefficient.

Table 7 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.

## 5.3. Interest personas are targeted personalized ads

Next, we analyze the ads delivered through `prebid.js`. 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 analyze thousands of ads. To this end, we sample a relatively manageable number of ads where we expect to see the most personalization.

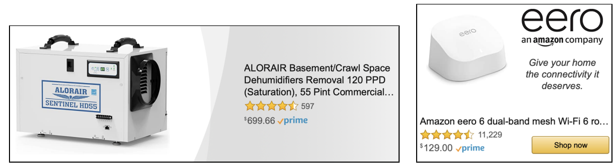
We consider an ad to be personalized if three conditions are met: *(i)* the skill vendor is also the advertiser (e.g., a Ford ad shown to a persona with “FordPass” skill), including Amazon itself, *(ii)* it is only present in one persona, and *(iii)* it references a product in the same industry as the installed skill, (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 and 255 ads from Amazon across 25 iterations. We manually inspect each ad and label them based on the text and product advertised in the ad.

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 the remaining 5, 3 are from Ford and 2 are from Jeep in Connected Car persona. It is noteworthy that 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.

| 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 8: Personalized 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.

However, ads from Amazon do seem to be personalized to personas. Table 8 presents the unique and personalized ads from Amazon. Health & Fitness and Smart Home personas receive unique and personalized ads, whereas Religion & Spirituality and Pets & Animals receive unique but non-personalized ads. The dehumidifier ad (Figure 4a) 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 [23] 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 [22]; both ads appeared once in Smart Home persona. We notice several ads repeated across iterations in Religion & Spirituality and Pets & Animals persona that do not seem to have any apparent personalization. For example, Amazon Eero WiFi (Figure 4b), Amazon Kindle, and Swarovski ads exclusively appeared 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.



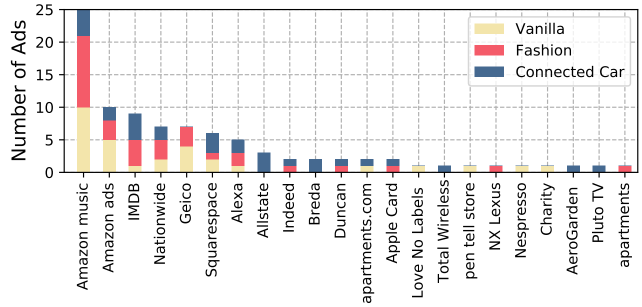

(a) Dehumidifier ad in Health & Fitness (b) Eero WiFi ad in Religion & Spirituality

Figure 4: Unique and repeated ads in interest personas.

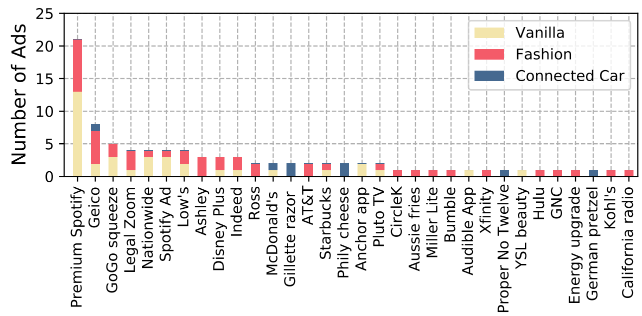
## 5.4. Audio ads are likely personalized

Next, we take a preliminary look at the 289 audio ads collected on Amazon Music, Spotify, and Pandora (Section 3.3). Table 9 shows the fraction of ads on each audio-streaming skill for each persona. Since the recorded audio for each skill is approximately equal in length, we surmise that differences in the number of ads streamed across personas on the same skill, signal differences in advertiser interest [57]. For instance, as shown in Table 9, 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.

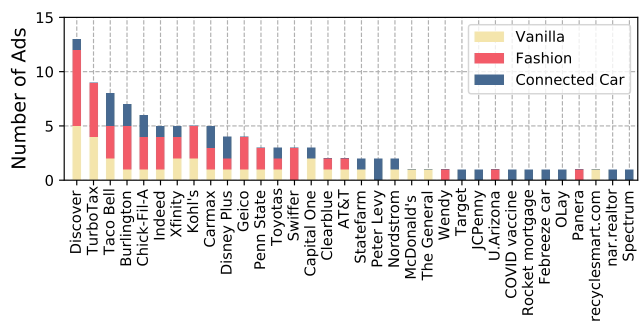
We also manually label the products advertised in order to look for evidence of obvious personalization (as we do in Section 5.3 for web ads). In this case, we only consider audio ads streamed twice or more, as repetitions may signal a stronger interest by the advertiser. Figure 5 present the distribution 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, see Section 5.3), 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.




(a) Audio ads on Amazon Music




(b) Audio ads on Spotify




(c) Audio ads on Pandora

Figure 5: Distribution of audio ads across Amazon Music, Spotify, and Pandora.

| **Persona** | **Amazon** | **Spotify** | **Pandora** |
|---|---|---|---|
| **Connected Car** | 33.33% | 8.99% | 26.17% |
| **Fashion & Style** | 34.41% | 50.56% | 43.92% |
| **Vanilla** | 32.26% | 40.45% | 29.91% |

TABLE 9: Fraction of ads ($n = 289$) on each audio-streaming skill for each persona.

### 5.5. 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 [55]. 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 does not syncs its cookies with any advertiser.[3] The one sided cookie-syncs could be explained by Amazon advertising's recent services for central identity resolution [86].

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 6 presents the bid values on common ad slots by Amazon's partners and non-partners 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.

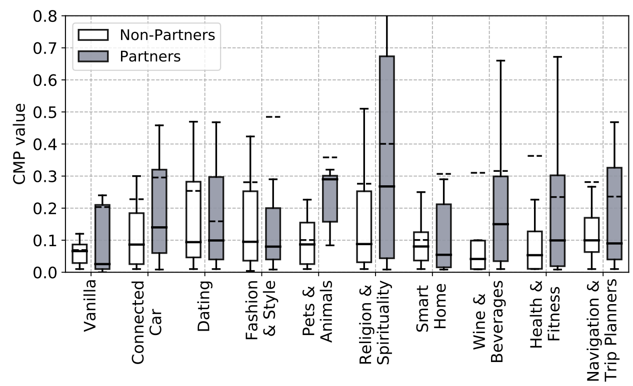



Figure 6: Bid values across personas on common ad slots distributed by Amazon's advertising partners.

### 5.6. Echo interest personas are targeted similar to web interest personas

This section expands the discussion we have in Section 5.6. 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

3. We analyze the OpenWPM datasets released by prior work [67] to validate that Amazon's cookie syncing behavior is not unique to our dataset.

| Persona | Partner Median | Partner Mean | Non-partner Median | Non-partner 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 10: Median and mean bid values for personas from Amazon's partner and non-partner advertisers.

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 11 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 | $p$-value Science | $p$-value 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 11: Statistical significance between Echo interest (persona column) and web interest (Health, Science, Computers columns) personas. $p$-value is computed through Mann-Whitney U test.

***Takeaway.*** Our measurements indicate the usage of voice data 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 Echo users. Some advertisers sync their cookies with Amazon and bid higher than non-cookie syncing advertisers.

## 6. Data Profiling Analysis

### 6.1. Amazon uses voice data to infer advertising interests

Since, Amazon allows users to access data collected about them, we request data for interest and vanilla personas [10]. The data contains detailed information about device

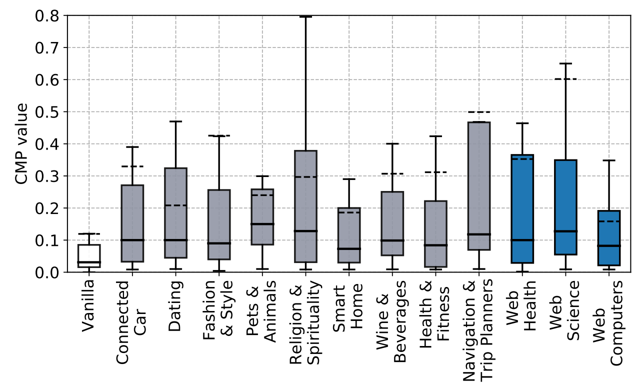



Figure 7: CPM values across vanilla, Echo interest, and web interest personas on common ad slots. Solid and dotted lines in bars represent median and mean, respectively.

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 requests data thrice, once after skill installation and twice after skill interaction. We request interest twice to see whether inferred interests evolve over time. Table 12 presents the advertising interests inferred by Amazon for various personas. 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 persona and also refine interests for Health & Fitness persona. Table 12 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 usage of data.

It is notable that the advertising interest inference that we observe can be interpreted as inconsistent with Amazon's public statements [83], [75]. Specifically, in a statement, Amazon mentioned that they do "not use voice recordings to target ads" [83], [75]. While Amazon may not literally be using the "recordings" (as opposed to transcripts and corresponding activities), our results suggest that they are processing voice recordings, inferring interests, and using those interests to target ads—this distinction between voice recordings and processed recordings may not be meaningful to many users. Amazon's policies state 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 [4]. The potential inconsistency between policies/statements and actual practices raises questions about Amazon's commitment to only using user data for stated purposes.

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

TABLE 12: Advertising interests inferred by Amazon for interest personas.

### 6.2. It is unclear whether skills play a role in targeting of personalized ads

Next, 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 1 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 5.3 and 6.1 already show that Amazon targets personalized ads to users and uses voice data to infer advertising interests, respectively. We also note that Smart Home, Wine & Beverages, and Navigation & Trip Planners, personas do not contact any non-Amazon services but still receives high bid values, as compared to vanilla persona. Amazon also infers discernible interests for the Smart Home persona (Table 12). 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 1 and Figure 2), which cannot lead to mass targeting. Similarly, we note that none of the skills re-target ads to personas (Section 5.3), 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.

***Takeaway.*** Amazon's inference of advertising interests from users' voice can be interpreted as inconsistent with their public statements. Amazon does not provide transparency in usage of data and thus cannot be reliably trusted to protect user privacy. Our findings indicate that skills require Amazon's collaboration to effectively use collected data.

## 7. Analyzing Privacy Policies

In this section, we analyze the consistency between the actual data collection practices and privacy policies.

### 7.1. Collecting Privacy Policies

First, we obtain the privacy policy of Amazon (platform) from its website [11]. This applies to all Amazon products, including Alexa. Alexa and its privacy controls are further described on the Alexa website [49]. We then download skills' policies using a Puppeteer [37] based crawler. We crawl the webpage of each skill, attempt to find the privacy policy link, and download it if there is one. Recall from Section 3.1.1 that we experiment with 450 skills: nine categories, top-50 skills per category. Among the 450 skills, only 214 (47.6%) skills provide links to their privacy policies and only 188 privacy policies can be downloaded. This is higher than the statistics reported by prior work [71], which identified that only 28.5% of the skills provide a privacy policy link [71]. Among the 188 obtained privacy policies, 129 do not even mention Alexa or Amazon in their text. They are mostly generic and apply to various products from the same developer—not specific to Alexa skills.

### 7.2. Network Traffic vs. Privacy Policies

We use and adapt PoliCheck [53] to perform NLP analysis of the privacy policies and to check the consistency of data flows found in the network traffic with those declared in the corresponding privacy policy. Policheck has been previously applied to mobile app traffic [53], traffic from VR headsets [84], as well as to voice assistants [71]. However, in [71], data flows were extracted not from actual network traffic (as we do in this paper), but from the permissions of skills [71].

In this context, a *data flow* is defined as *<data type, endpoint>*, i.e., what data type is sent to what destination endpoint (or "entity" in Policheck terminology [53]). While running a skill, PoliCheck (i) extracts data flows as *<data type, entity>* tuples from the network traffic of the AVS Echo (ii) analyzes the corresponding skill's privacy

| 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 13: 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 .

policy text for statements that disclose these data flows and (iii) checks the consistency of the two. For example, while running the skill *Sonos* [44], the AVS Echo's network traffic includes an outgoing packet that sends voice data to an Amazon endpoint; PoliCheck will extract the tuple <*voice, amazon*> from this packet. At the same time, Sonos states the following text in its privacy policy: *"The actual recording of your voice command is then sent to the voice partner you have authorized to receive such recording (for example, Amazon)."* Thus, PoliCheck will also extract the tuple <*voice, amazon*> from this statement. Since the tuple from the network traffic matches the statement tuple in the privacy policy, PoliCheck labels this as a *clear* disclosure. In general, a data flow found in the network traffic can be classified by Policheck [53] as: clear, vague, ambiguous, incorrect, or omitted.

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 does not support certain features and only communicates with Amazon endpoints), 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. Note that we have adapted two distinct versions of PoliCheck based on the version released in [84] to perform these two analyses separately, as described next.

**7.2.1. Endpoint analysis.** Since the encrypted traffic does not reveal the exact data types, we modify PoliCheck to focus on validating entities (i.e., names of organizations) during endpoint analysis. We update PoliCheck's entity ontology to include all the 13 endpoints we observe—each endpoint organization is labeled with one or more categories: *analytic provider*, *advertising network*, and *content provider* (see Table 14). Amazon, as platform-party, is also labeled as *platform provider* and *voice assistant service*. Next, we classify endpoint consistency into one of three disclosure types: (1) *clear*, when the endpoint is 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 as in the original PoliCheck because a contradiction cannot be determined without considering data types. Finally, we label an endpoint as (4) *no policy* when the skill does not provide a privacy policy. Table 14 presents the result of our endpoint analysis.

*Disclosure of platform-party collection.* Only 10 privacy policies clearly indicate the possibility that personal information is collected by Amazon. For example, the skill *Sonos* [44] clearly states that voice recording is collected by Amazon. Furthermore, we also found 136 skills, whose statements contain vague disclosures that may correspond to the traffic going to Amazon. For example, the privacy policy of the skill *Harmony* [26]) has the following statement, in which Amazon is not explicitly mentioned as an entity that: *"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 found 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 skills *YouVersion Bible* [50] and *Garmin* [24], and correspond to the organizations that are the developers of the skills.

*Disclosure of third-party collection.* Many third-party endpoints, e.g., Liberated Syndication, Podtrac, Spotify and Triton Digital, provide audio content distribution and monetization (tracking/advertising) services. Skills likely rely on these third-party service providers to deliver audio contents. However, only a few skills disclose data collection and sharing with third-party endpoints in their privacy policies, and when they do, they use vague terms. For example, the skill *Charles Stanley Radio* [17] uses the term *"external service providers"* to refer to third-party endpoints in the following statement in its privacy policy: *"We may also share your personal information with external service providers who help us better serve you."* Another example is the skill *VCA Animal Hospitals* that uses the blanket term *"third-parties"* to refer to all third-party endpoints in its privacy policy [48].

**7.2.2. Data Types Analysis.** We adapt PoliCheck to perform consistency analysis on the data types found in the unencrypted traffic collected from the AVS Echo. Thus, we rebuild PoliCheck's data ontology by following the methodology used in previous work [71], [84]. We add new terms that represent new data types, particularly *voice recording* that is unique to voice assistants. Furthermore, we also improve this version of PoliCheck: we modify it to focus on checking specific data types and ignore vague terms, e.g., *pii*, *user info*, and *technical info*. Finally, we classify data types consistency using the same disclosure types used in endpoint analysis. Table 13 presents the result of our data types analysis using PoliCheck.

*Disclosure of data types in skill's privacy policies.* 83 skills have at least one clear or vague disclosures. Among them, only 20 and 11 skills disclose the collection of voice recordings and customer IDs clearly. Finally, despite providing privacy policies, 174 skills do not disclose the collection of data types observable in their network traffic.

*Disclosure of data types in Amazon's privacy policy.* As noted in Section 7.1, only 59 skills mention Amazon or Alexa in their privacy policies. Among these, only 10 of them explicitly provide a link to Amazon's privacy policy or terms of use. In addition to the low availability and specificity of skills' privacy policies, we identify a gap between developers and Amazon: most developers neither disclose the data types in their privacy policies nor provide a link to Amazon's privacy policy, possibly because they are not aware that Amazon is collecting these data types when a skill is running.

Going forward, we believe that the good practice of a developer referencing the platform's privacy policy in the skill's privacy policy is easy to adopt. *What would be the impact of this practice to the clarity of disclosures?* Following the methodology in [84], we set PoliCheck to also check the platform-party's privacy policy, by default, and we perform another experiment: we include Amazon's privacy policy in addition to the skill's own privacy policy. We find that PoliCheck classifies all data flows to be either clearly or vaguely disclosed depending on the terms that Amazon's privacy policy uses to disclose the data types. Table 13 lists the terms found in the Amazon's privacy policy by PoliCheck.

***Takeaway.*** In general, our findings suggest that the majority of skill developers, even among the top skills, do not write their privacy policies properly. In other words, the skills' actual data collection and sharing practices are often not clearly disclosed in their privacy policies.

**7.2.3. Validation of PoliCheck results.** To validate the correctness of PoliCheck when applied to skills, we visually 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 [84], [71], [53], we consider *multi-class classification*. Similarly to [84], 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.

## 8. 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 used Amazon Echo interactions 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.

### 8.1. 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 [47]. Having such interface unlocked for developers and auditors would reveal the actual data being shared. Another example of a possible user defense is

| 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 selectively block network traffic that is not essential for the skill to work (e.g., using an approach similar to [72]).

*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. [82]). 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 [35], [39], and just send to the Alexa platform the transcribed commands using their textual API with no loss of functionality.

## 8.2. Parallels with Other IoT Platforms

*Related platform-agnostic IoT works.* Several IoT works have measured network traffic to detect data sharing. For example, [73], [65], [79], [80], [72] 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 [72], 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 1) and 12 new third-party endpoints (see *third party* row in Table 1). A possible explanation could be the change in Amazon's ecosystem since it was last studies, 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 [85], [74] and VR headsets [84], 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.

### 8.3. Clarifications and Updates

Since the initial release of this paper on arXiv [66], we have updated it to clarify some statements, so as to avoid possible misinterpretations. In particular, we do not claim that Amazon directly shares voice recordings or transcripts with advertising networks. Neither do we claim that Amazon surreptitiously records users' voices; we issued voice commands and expected to be recorded. We do find evidence that Amazon processes voice recordings from skill interactions to infer user interests; and that it uses those interests to target ads. We also clarified that Amazon's inference of advertising interests from users' voice is potentially inconsistent with their public statements. Amazon's and skills' operational practices are often not clearly disclosed in their privacy policies. Amazon's privacy policy neither acknowledges nor denies the usage of Echo interactions for ad targeting. We have also made more precise claims regarding Amazon's advertising partners syncing their cookies with Amazon, avoiding language specifying that Amazon shares user interests with advertisers.

## References


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

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

[3] 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.

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

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

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

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

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

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

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

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

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

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

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

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

[16] 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.

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

[18] 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.

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

[20] 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.

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

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

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

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

[25] 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.

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

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

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

[29] 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.

[30] 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.

[31] 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.

[32] 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.

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

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

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

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

[37] Puppeteer. https://www.npmjs.com/package/puppeteer.

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

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

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

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

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

[43] 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/.

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

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

[46] 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).

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

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

[49] You have control over your Alexa experience. https://www.amazon.com/b/?node=19149155011.

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

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

[52] 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/.

[53] 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.

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

[55] 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).

[56] 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).

[57] 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).

[58] 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).

[59] 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).

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

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

[62] 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.

[63] 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/.

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

[65] 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.

[66] IQBAL, U., BAHRAMI, P. N., TRIMANANDA, R., CUI, H., GAMERO-GARRIDO, A., DUBOIS, D., CHOFFNES, D., MARKOPOULOU, A., ROESNER, F., AND SHAFIQ, Z. Your echos are heard: Tracking, profiling, and ad targeting in the amazon smart speaker ecosystem. *arXiv 2204.10920v1* (April 22, 2022).

[67] 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).

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

[69] JIN, H., AND WANG, S. Voice-based determination of physical and emotional characteristics of users, Oct. 2018. US Patent 10,096,319.

[70] 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.

[71] 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).

[72] 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).

[73] 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.

[74] 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.

[75] 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/.

[76] OLEJNIK, L., TRAN, M.-D., AND CASTELLUCCIA, C. Selling off privacy at auction. In *Network and Distributed System Security Symposium (NDSS)* (2014).

[77] 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).

[78] PRO, A. P. Adobe premiere pro, 2022.

[79] 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).

[80] 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).

[81] 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.

[82] SINGH, R. *Profiling humans from their voice*. Springer, 2019.

[83] 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.

[84] 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).

[85] 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).

[86] 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.

[87] 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).