Introduction

The main goal of the Voice Assistant Profiling project was to investigate whether Amazon profiles users surreptitiously when using their voice assistant (i.e. Alexa) and to determine how this is represented if any profiling is found. To identify how Amazon profiles their users, we analyzed the "Amazon Audiences" that appear in the Advertising folder when requesting for data. Amazon Audiences refer to interests used to identify what categories of products a user would be interested in — an example could be "In-market: Toys and Games" to denote a user who is interested in purchasing such items. These Audiences are used to suggest targeted ads according to interests – for example, those who have the "In-Market: Toys and Games" label may more frequently notice product recommendations for toys.

Methodology

We created 2 personas (Video Entertainment and Fashion) and aimed to search up queries that we thought were likely to produce those Audiences. The Video Entertainment persona was created to resemble a user interested in movies, TV shows, etc., while the Fashion persona was created to resemble a user interested in clothes, accessories, shoes, etc. For each persona, we curated a list of 20 questions to fulfill the targeted interests.

To curate the questions, we listed keywords that were likely to be associated with a certain profile. For example, to target the Video Entertainment audience, we would search for queries related to movie rentals, ratings on IMDB, etc. Likewise, for the Fashion audience, we would search for queries related to articles of clothing.

To complete the voice searches, we recorded ourselves enunciating the queries and played them back to the voice assistant device. This was done to maintain the pitch, tone, emotion, etc. of our voices so that each iteration of our experiment was consistent. Furthermore, this limited the possibility of Amazon profiling our accounts differently if characteristics of a person's voice were taken into account in the profiling process.

Upon completing the voice searches, there would be a waiting period of 1 week to allow Amazon to complete the profiling process. Then, we requested the Advertising data from Amazon's servers (My Account -> Request Your Data -> Advertising) and allow a few hours for the request to be completed.

When requesting the Advertising data, we found that Amazon will return 3 folders labeled "Advertising.1", "Advertising.2", and "Advertising.3" — we focused on the contents of "Advertising.1" as we believe that the other 2 folders were specific to different regions where Amazon is also available (in this case, Europe and Asia). Within the "Advertising.1" folder, we analyzed the CSV file titled "Advertising.AmazonAudiences", which shows what categories of products Amazon believes you are interested in.

Results

Newly created accounts will not produce an "Advertising.AmazonAudiences" file upon requesting data, and this file will only show up after a sufficient number of queries have been searched up. We noted that these queries also have to be related to products that Amazon offer (e.g. "Search for cheap black dresses"), and that more general queries (e.g. "Suggest restaurants near me") do not prompt a profiling process from Amazon.

Interestingly, there seems to be a discrepancy between the availability of certain files when requesting data: as shown in Figure 1 below, 1 file appears in the "Advertising.1" folder on the 1st day; 2 files appear on the 4th day; 1 file appears on the 6th day; 3 files appear on the 7th day.

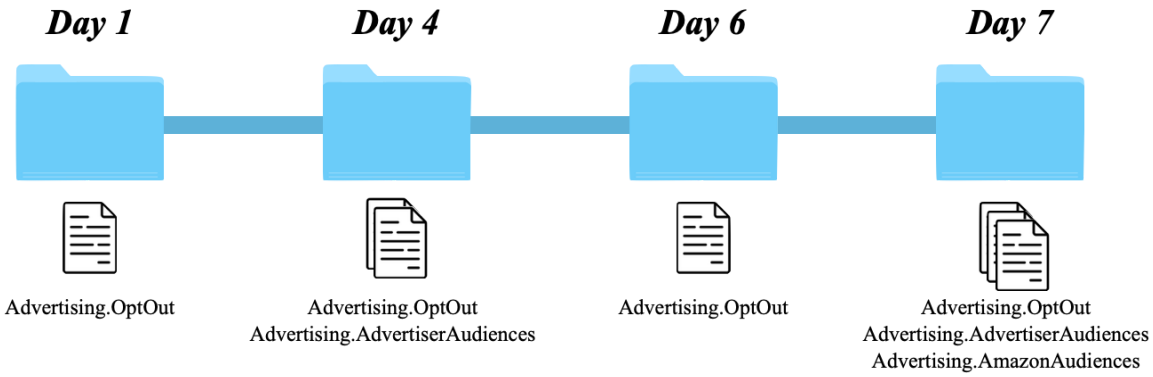

Figure 1

However, we found that Amazon has remained consistent in other regards of profiling — Amazon has profiled each instance of a persona in a similar manner as its counterparts. This is illustrated in the tables below:

VIDEO ENTERTAINMENT PERSONA		
	Target	**Result**
Account 1	• In-market: Video Entertainment	• In-market: Video Entertainment • In-market: Video Entertainment: Prime Video
Account 2	• In-market: Video Entertainment	• In-market: Video Entertainment • In-market: Video Entertainment: Prime Video

Account 3	• In-market: Video Entertainment	• In-market: Video Entertainment • In-market: Video Entertainment: Prime Video
Account 4	• In-market: Video Entertainment	• In-market: Video Entertainment • In-market: Video Entertainment: Prime Video

FASHION PERSONA		
	Target	**Result**
Account 1	• In-market: Fashion	• In-market: Fashion • In-market: Beauty & Personal Care
Account 2	• In-market: Fashion	• In-market: Fashion • In-market: Beauty & Personal Care
Account 3	• In-market: Fashion	• In-market: Fashion • In-market: Beauty & Personal Care
Account 4	• In-market: Fashion	• In-market: Fashion • In-market: Beauty & Personal Care

Through probing questions (i.e. neutral queries that should not produce a biased result), which were searched up after each account has been profiled, we found that the results Alexa presented were consistent across both personas — for example, the query "Suggest a restaurant near Harvard" will produce an identical list for both the Video Entertainment and Fashion persona, which led us to believe that Amazon is not secretly imposing a bias towards users with certain interests.

We also found that the "Advertising.AdvertiserAudiences" file produced the same list of Advertiser Audiences across the 2 personas, despite the difference in Amazon Audiences. We speculate that only a certain number of Amazon Audiences would cause a change and/or addition to the list of Advertiser Audiences.

Amazon vs. Google

It is clear that Amazon's profiling system is more rudimentary compared to Google's — while Google assigns "tags" to encapsulate a user's demographics and includes a more comprehensive list of Ad Topics (i.e. a user's interests for advertisers to push their content), we found that Amazon only profiles users based on their interests in product categories, but not in their demographics such as age, economic class, gender, etc.

We also found that Amazon adheres to a user's choice of location privacy settings. If a user chooses not to share their location, queries such as "Search for restaurants **near me**", Amazon will produce an invalid response to note that it is unable to search for spots within your vicinity. Google, on the other hand, seems to be able to locate a user's approximate position despite not having access to location services.

Lastly, Amazon does not allow personalization of profiling. While Google allows users to correct their "tags" (e.g. change their tag from "Homeowners" to "Renters) and remove Ad Topics that they are not interested in, there does not seem to be a way to do so for Amazon. This means that if Amazon profiled a user incorrectly, they are unable to rectify the inconsistency and are forced to be presented with possibly inaccurate ads and product recommendations.

Challenges

Throughout our experiments, the main challenge we ran into was creating an adequate number of accounts to perform multiple experiments on, as Amazon started prompting for a phone number upon account creation. To overcome this, we had to wait around 1 week before Amazon whitelisted our devices to allow for more account creation. Secondly, we found that automation was not much of an option when conducting the voice searches. As the Alexa app would occasionally jump to the Amazon and Prime Video app when processing a query, meaning that the Alexa app was not always active, there was no way to play all the voice recordings in one iteration. However, due to the small number of queries, each experiment would take 5-10 minutes to complete. Lastly, we found that our initial list of queries, which were more general-purpose than product-specific, did not trigger a profiling process. Hence, we had to refine our queries to target Amazon's version of profiling.

Conclusion

In summary, we have suggested evidence that Amazon does profile its users based on product categories, but are not doing so surreptitiously and in a way that produces biased results when searching up probing questions. However, we believe that Amazon should allow for more customization to allow users to opt in/out of certain product categories. Allowing advertisers to continue pushing ads to certain users who are not actually interested in their products may also be disadvantageous and less profitable for advertisers.

There is more work to be done to further validate our results, and if given the opportunity and time, we would like to determine if certain Amazon audiences are more likely to introduce new Advertiser Audiences to the list.