Mitigating Security Risks in Commonly Used Alexa Skills

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Jammie Wang
Spring, 2021

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____Jammie Wang________________ Date __5/4/2021___
Jammie Wang

Approved ____ Date __________
Yuan Tian, Department of Computer Science

Abstract

With the growing popularity of voice assistants, privacy has become a key issue. Amazon Alexa skills have had privacy issues in the past, yet the device is only becoming more popular. Using a dataset of critical reviews, this research study aims to identify common “risky” categories of Alexa skills, particularly those that exhibit privacy and security concerns. The frequency of these privacy concerns will also be evaluated in comparison to the frequency of other types of concerns, namely financial concerns or concerns about inappropriate content. Each review in the data set is parsed for a set of “keywords” that defines a specified issue.

Most mentions of inappropriate content are from the “Games & Trivia” category of skills. Financial complaints were found to be much more common across all categories. 38% of these finance-related reviews belonged in the “Games & Trivia” category. Privacy and security concerns were present across all categories of Alexa skills in this data set, though less frequently than financial concerns. The “Kids” category appears to have a handful of skills that are particularly unsafe compared to the rest. Privacy concerns are more common in “Social”, “Kids”, and “Food & Drink” categories than financial issues. These results indicate that Alexa may need additional content moderation and stricter privacy policies to prevent skills from targeting vulnerable populations, such as children, for personal information.

Mitigating Security Risks in Commonly Used Alexa Skills

Introduction

Commercially available voice assistants on the market often exhibit a variety of privacy issues ranging from recording personal conversations to downloading skills that ask for extensive personal information (Pal et al., 2020). Amazon Echo (Alexa) faced public scrutiny in 2018 after it recorded a private conversation and sent it to another person on the owner's contact list (Sacks, 2018). Such incidences led to an extensive debate on the privacy of voice assistants, especially since 15.4% of the US population already owned an Amazon Echo in 2018 (Pridmore & Mols, 2020). With the growing use of voice assistants, privacy is becoming increasingly important.

Relevant Works

Current state of the art research has identified the relationship between the usage of voice assistants and its role in surveillance capitalism (Pridmore & Mols, 2020). Similar research has developed privacy-preserving trust models, which emphasizes a "privacy by design" approach for enhancing end-user trust (Pal et al., 2020). However, much of the current research does not account for categorical differences between privacy concerns. Categorical differences may mean that users may be more distrustful of certain categories, such as "Social" skills, or that some categories may have higher rates of privacy violations. Such differences imply that some categories require a certain approach to achieve better privacy and higher security, while other categories may require a different approach.

Purpose

This research study aims to identify common "risky" categories of Alexa skills, particularly those that exhibit privacy and security concerns. The frequency of these privacy concerns will also be evaluated in comparison to the frequency of other types of concerns, namely financial concerns or concerns about inappropriate content. We will also briefly address potential solutions to these privacy concerns. This expands on existing research by identifying "target" categories that either need more guidelines, or potentially stricter guidelines overall to ensure user privacy.

Research Methods

Dataset Structure

The dataset used for analysis contains two parts:

- A folder that contains 2973 JSON files. Each JSON file lists all the negative reviews for one particular Alexa skill. Some files may only contain one negative review, while others may have hundreds. The name of each JSON file is the skill ID that corresponds to the Alexa skill.
- A folder that contains roughly 54,000 skill profiles. Each skill profile includes information about the skill, such as the category it belongs to, and the skill description. Since there are less critical reviews in this dataset than skill profiles, not all skill profiles will be used during analysis. This dataset contains reviews from 10 categories. The name of each JSON file is once again the skill ID.

There are 25,233 negative reviews in the dataset, which spans across 10 categories. A large majority of these reviews belong to skills in the "Games & Trivia" category. Figure 1 shows the number of reviews in each category, and the corresponding percentage. The percentage is calculated by dividing the total number of reviews by the number of reviews within the respective category.

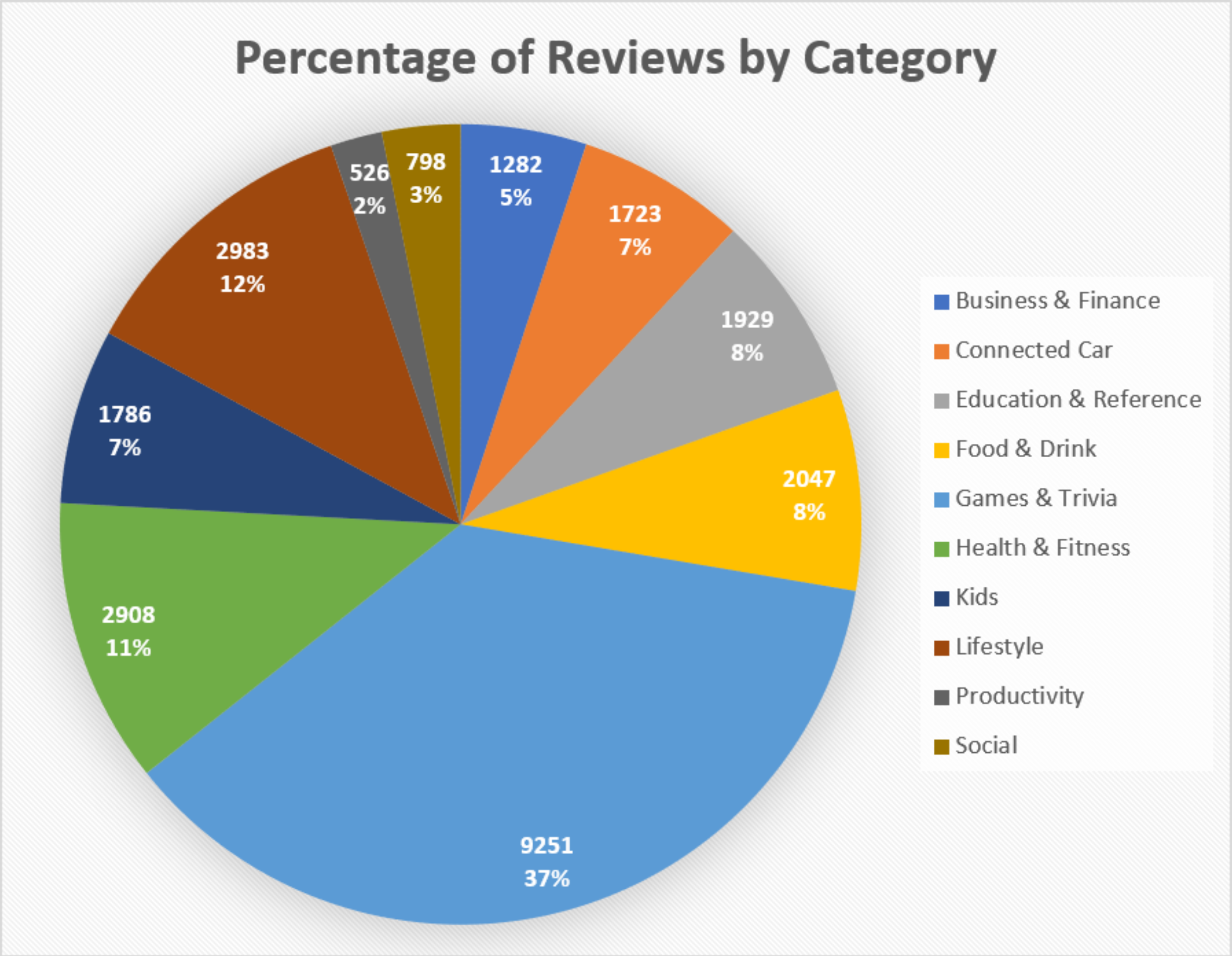


Figure 1: Distribution of reviews by skill category

As mentioned before, the format of the JSON files vary between Alexa skills. Some skills may only contain one review, while others contain multiple. A category may only have 5 skills, but hundreds of reviews for each skill in the category, while another category may have a hundred skills with one review each. Thus, the distribution of the number of skills in each category is different from the distribution of the number of reviews, though they are similar. The dataset contains 2973 skills between the 10 categories, which is illustrated in Figure 2. The percentages in Figure 2 is calculated by dividing the total number of skills by the number of skills within the respective category.

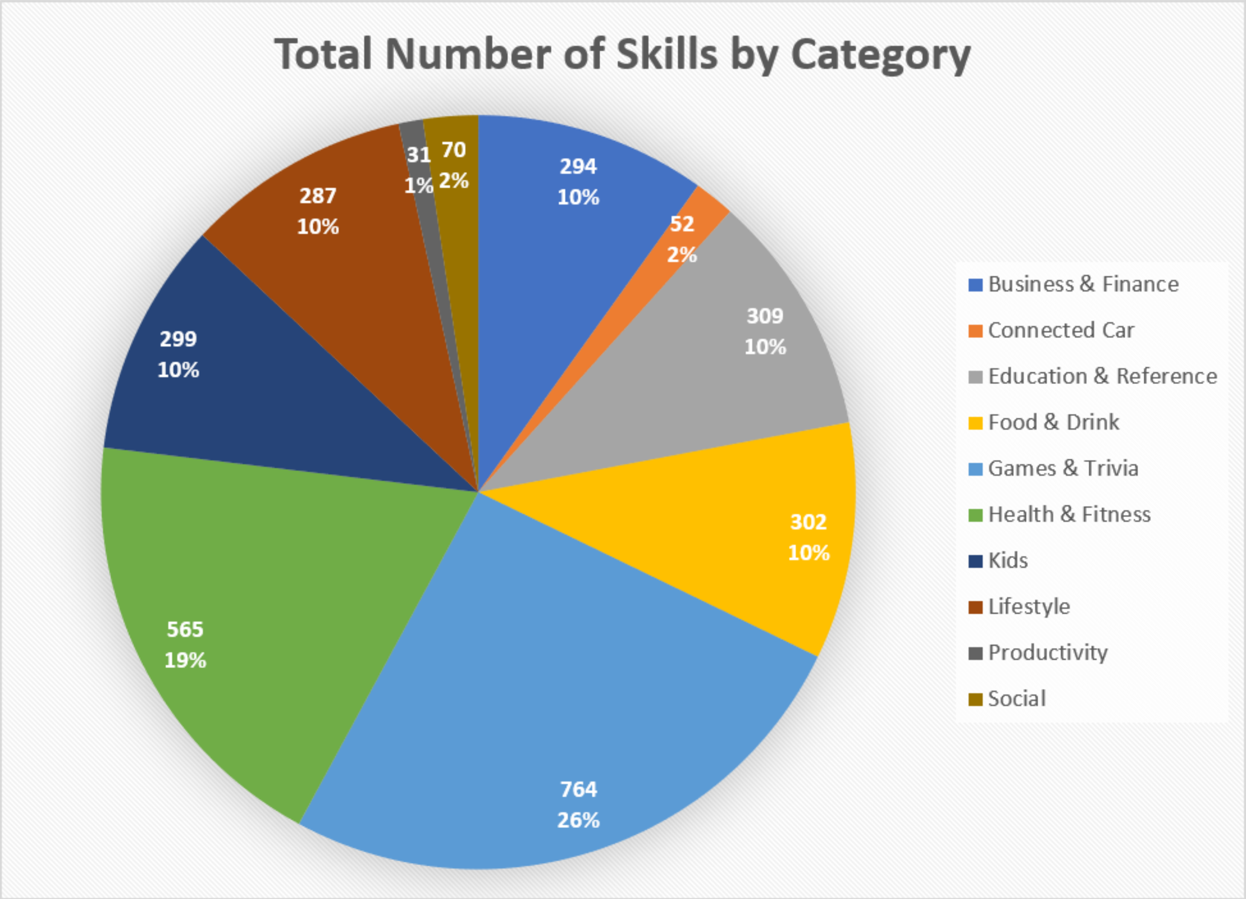


Figure 2: Distribution of skills by category

Algorithm

The goal of the algorithm is to count the number of occurrences of specific keywords given a dataset of negative Alexa reviews. These keywords are "negative words" commonly used by reviewers to express a concern or complaint about the Alexa skill. The algorithm operates as follows:

1. Set up a pandas dataframe such that each row is a skill category, and each column is a count of the number of specified keywords that have appeared in each category.
2. Loop through each JSON file in the folder of negative reviews. Extract the body of each review, and the skill ID for the Alexa skill.

3. Find the skill profile in the skill profiles folder using the skill ID. Extract the category the skill belongs to.
4. Loop through the body of each review for the skill. If a negative keyword appears in the review, update the dataframe by incrementing the keyword count of the appropriate category and type of concern.
5. Repeats steps 2, 3, and 4 until the algorithm has finished parsing all critical reviews in the critical reviews folder.

Limitations

This algorithm has several limitations in terms of effectiveness and quality of evaluation. Variations in natural language, including using synonyms, sentence structure, word misspellings, etc. are common in reviews. Reviewers may express a complaint without actually using one of the specified keywords. False positives may also be common in some cases. For example, a reviewer may casually mention that updating their personal information doesn't work. However, this does not mean that they have a privacy concern regarding their personal information. The algorithm does not detect such errors, which may limit the accuracy of the results to some extent.

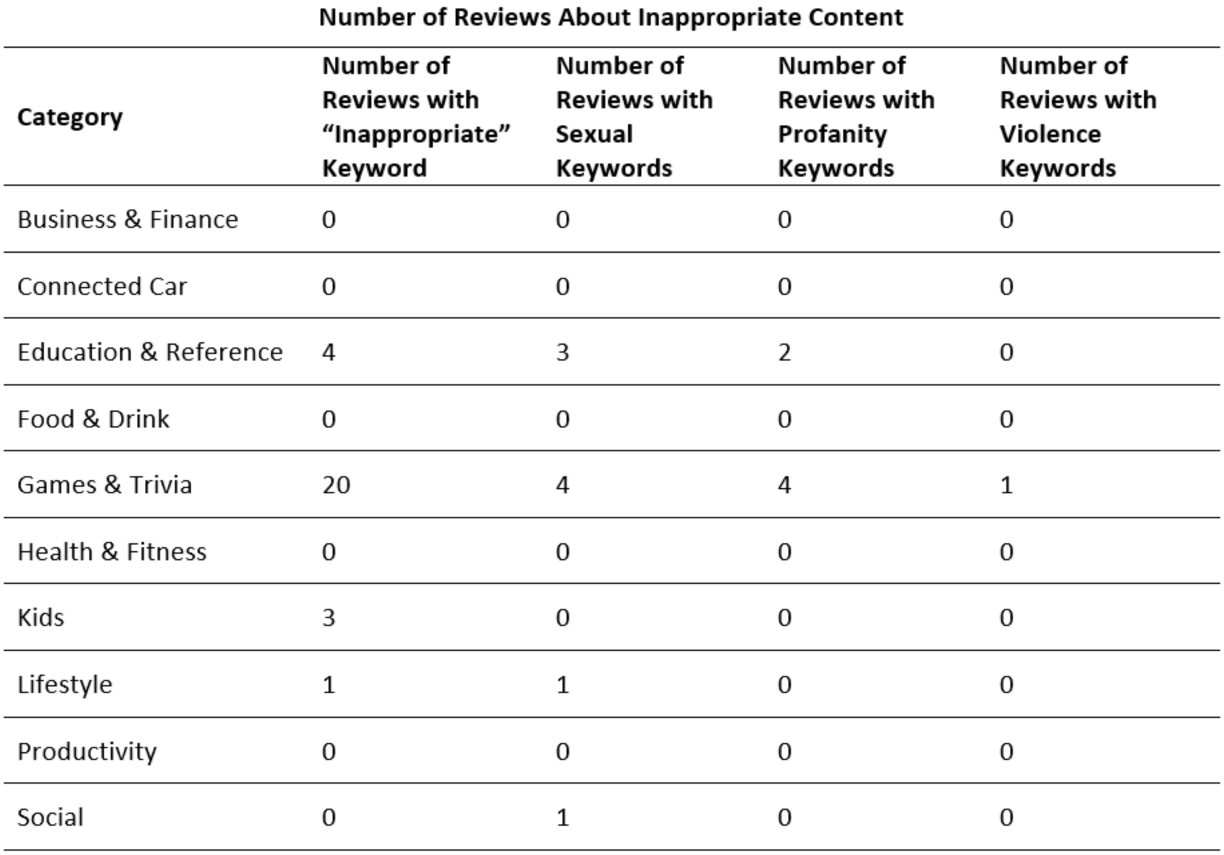
Results/Discussion

Inappropriate Content

One of the primary goals of this research is to identify issues users typically complained about. Inappropriate content was occasionally mentioned as a complaint. The number of reviews that contained the keyword "inappropriate" was counted. Most of the instances of the "inappropriate" keyword appeared in skills that belonged to the "Game & Trivia" category, as seen in Figure 5. To gain a better understanding of the types of inappropriate content that may be present, keywords may be used to detect sexual content, profanity, and violence. The keywords used to determine the type of inappropriate content are as follows:

- Sexual Content Keywords - ‘sex’, ‘sexual’
- Profanity Keywords - ‘profane’, 'profanity', 'cursing', 'swearing', 'f-word'
- Violence Keywords - 'violent', 'violence'

The frequency of these complaints are shown in Figure 3.



Number of Reviews About Inappropriate Content

Category	**Number of Reviews with “Inappropriate” Keyword**	**Number of Reviews with Sexual Keywords**	**Number of Reviews with Profanity Keywords**	**Number of Reviews with Violence Keywords**
Business & Finance	0	0	0	0
Connected Car	0	0	0	0
Education & Reference	4	3	2	0
Food & Drink	0	0	0	0
Games & Trivia	20	4	4	1
Health & Fitness	0	0	0	0
Kids	3	0	0	0
Lifestyle	1	1	0	0
Productivity	0	0	0	0
Social	0	1	0	0

Figure 3: Number of reviews mentioning an inappropriate keyword by category

Financial Complaints

Financial complaints were a major issue among reviewers for Alexa skills. Many complaints involved issues with fees, charges on credit cards, or overly expensive skills that reviewers felt were not worth the price. The keywords used to determine finance-related complaints are as follows: 'charged', 'fees', 'money', 'payment', 'pay'. The number of reviews related to financial complaints were much higher than the number of complaints about inappropriate content or privacy concerns. The number of complaints related to financial issues is shown in Figure 4, along with the corresponding percentage. The percentage is calculated by dividing the total number of finance-related reviews by the number of reviews in that category.

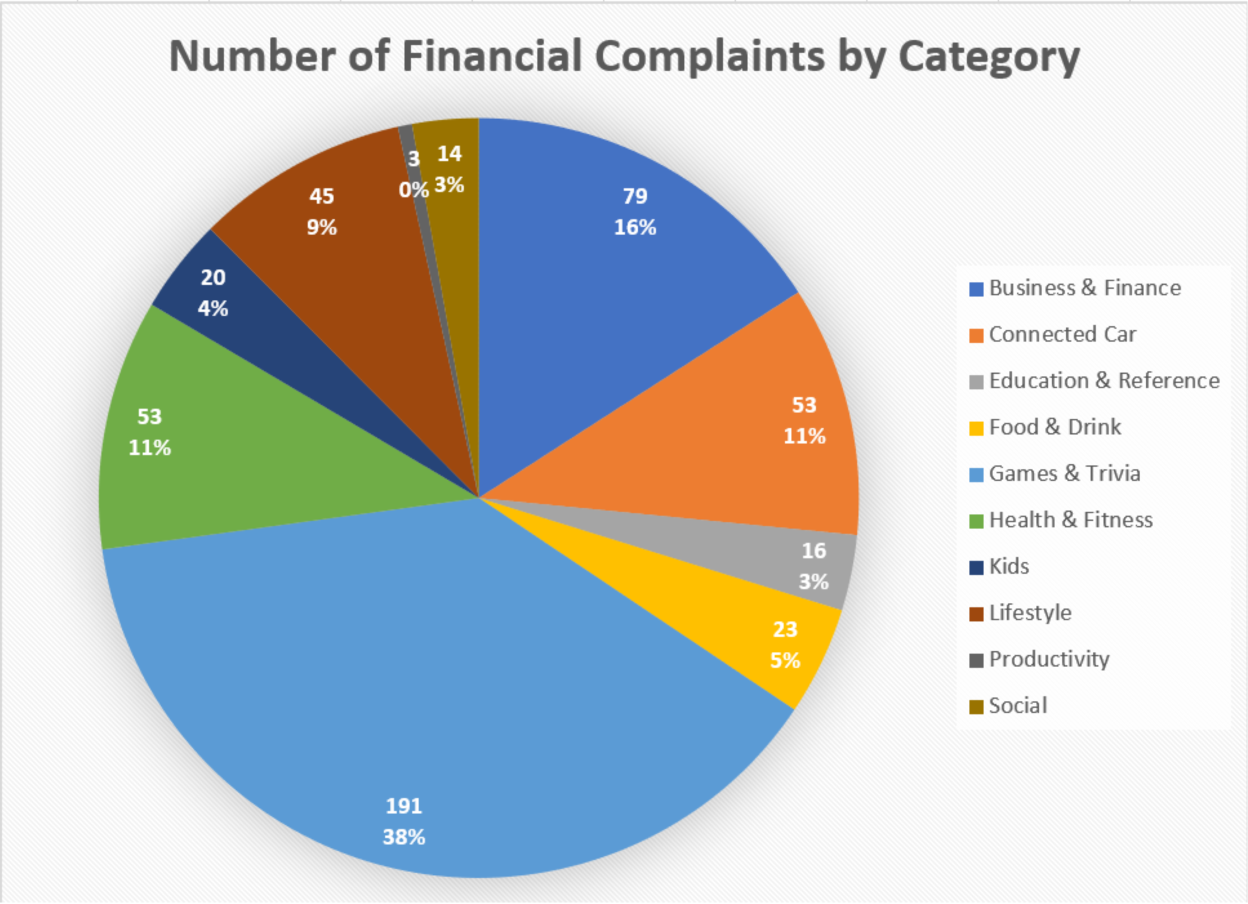


Figure 4: Number of reviews mentioning a financial issue by category and corresponding percentage

It is important to note that there is a higher chance of false positives in the Business & Finance category, as discussed before. Keywords such as ‘payment’ or ‘pay’ may be registered as a financial complaint when in reality, the complaint may have been the customer had issues paying a balance on their credit card through a banking Alexa skill.

Privacy Concerns

Like the finance-related complaints, reviewers expressed concern about privacy and security related issues throughout all categories of examined Alexa skills. The two ways this can be examined are through the number of reviews in each category that have privacy complaints, and through the number of skills that have privacy complaints in each category. Examining the data through both lenses is necessary because one particular skill could account for nearly all privacy concerns in a category, which would not be seen by examining the overall number of reviews in a category alone.

Privacy Review Distribution

The keywords used to identify privacy or security concerns are as follows: 'privacy', 'private', 'collect', 'personal', 'scam', 'collecting', 'collected', 'security', and 'secure'. Using these defined keywords, the results show that privacy concerns are prevalent to varying degrees in all examined categories, as illustrated in Figure 5. There are a total 194 reviews that exhibit security or privacy concerns.

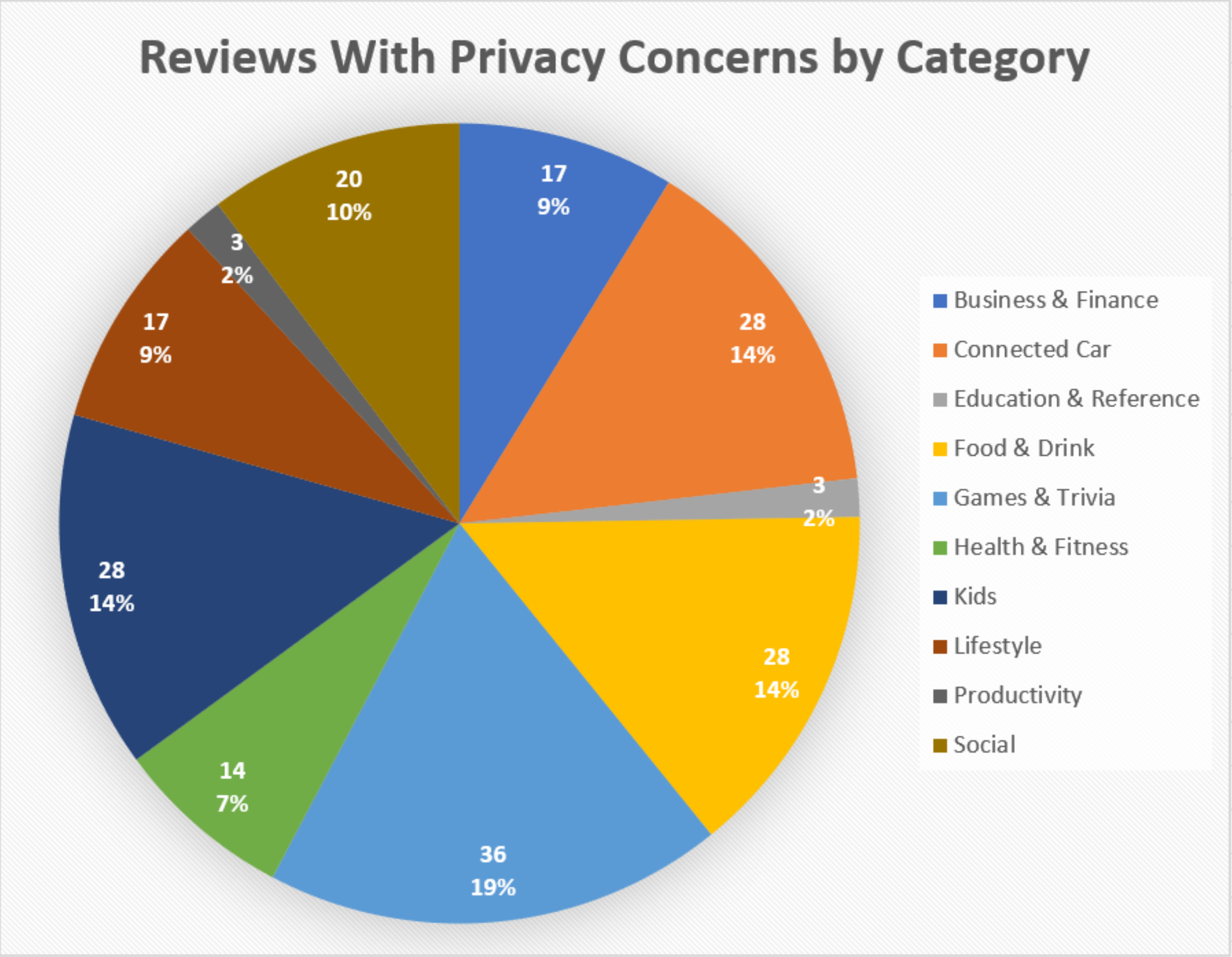


Figure 5: Number of reviews with privacy complaints and their percentages

Despite the "Games & Trivia" category accounting for nearly 37% of the total number of reviews, this category accounted for only 19% of the privacy-related reviews. Similarly, the "Health & Fitness" category accounted for about 11% of the total reviews, yet accounted for only 7% of the total number of reviews with privacy concerns.

There are also categories that show the opposite trend. Some categories hold a larger percentage of reviews with privacy concerns, compared to their percentage in the total reviews. Such categories include "Business & Finance", "Connected Car", "Food & Drink", "Kids", and "Productivity." It is important to note that since the frequency of privacy concerns were relatively low, a possible false positive or false negative would have a large impact on the percentage calculations.

Problematic Skills Distribution

In the former section, the number of reviews with privacy complaints are the primary evidence of larger categorical trends. To ensure that there are not cases of one skill having excessive privacy complaints and skewing the data for an entire category, it's essential to look at the distribution of the number of skills with privacy complaints. As seen in Figure 6, the distribution of problematic skills is fairly similar to the distribution of negative reviews.

However, two categories have noteworthy differences between the two distributions: "Games & Trivia" and "Kids." The "Games & Trivia" category accounted for 19% of the reviews with privacy concerns, yet accounted for 26% of the skills with privacy concerns. This suggests that privacy concerns are more widespread in the "Games & Trivia" category than it originally appeared. On the other hand, the "Kids" category accounted for 14% of reviews with privacy concerns, yet only account for 8% of skills with privacy concerns. This suggests that there are a handful of skills that are particularly unsafe in the "Kids" category, and account for many of the negative privacy-related reviews.

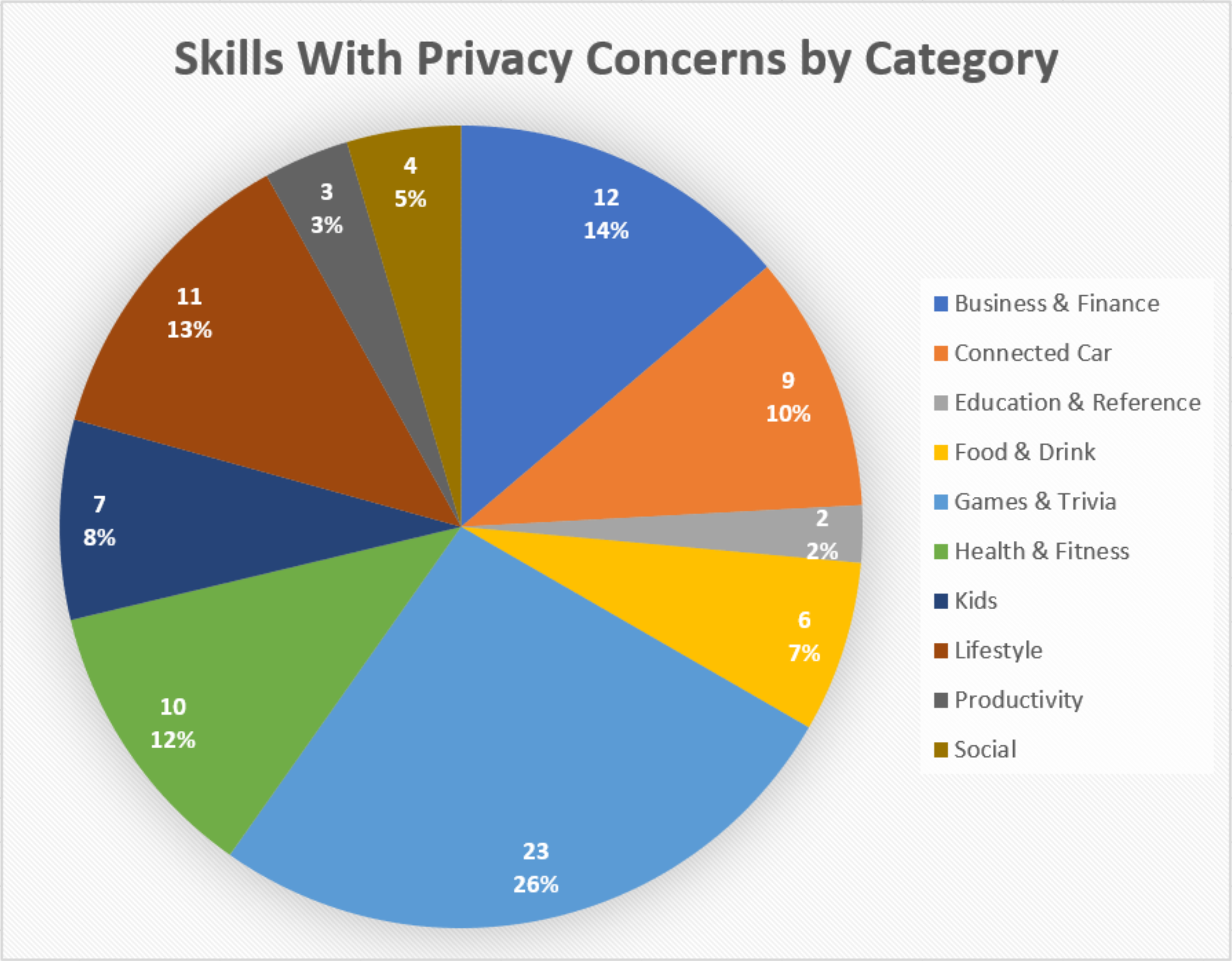
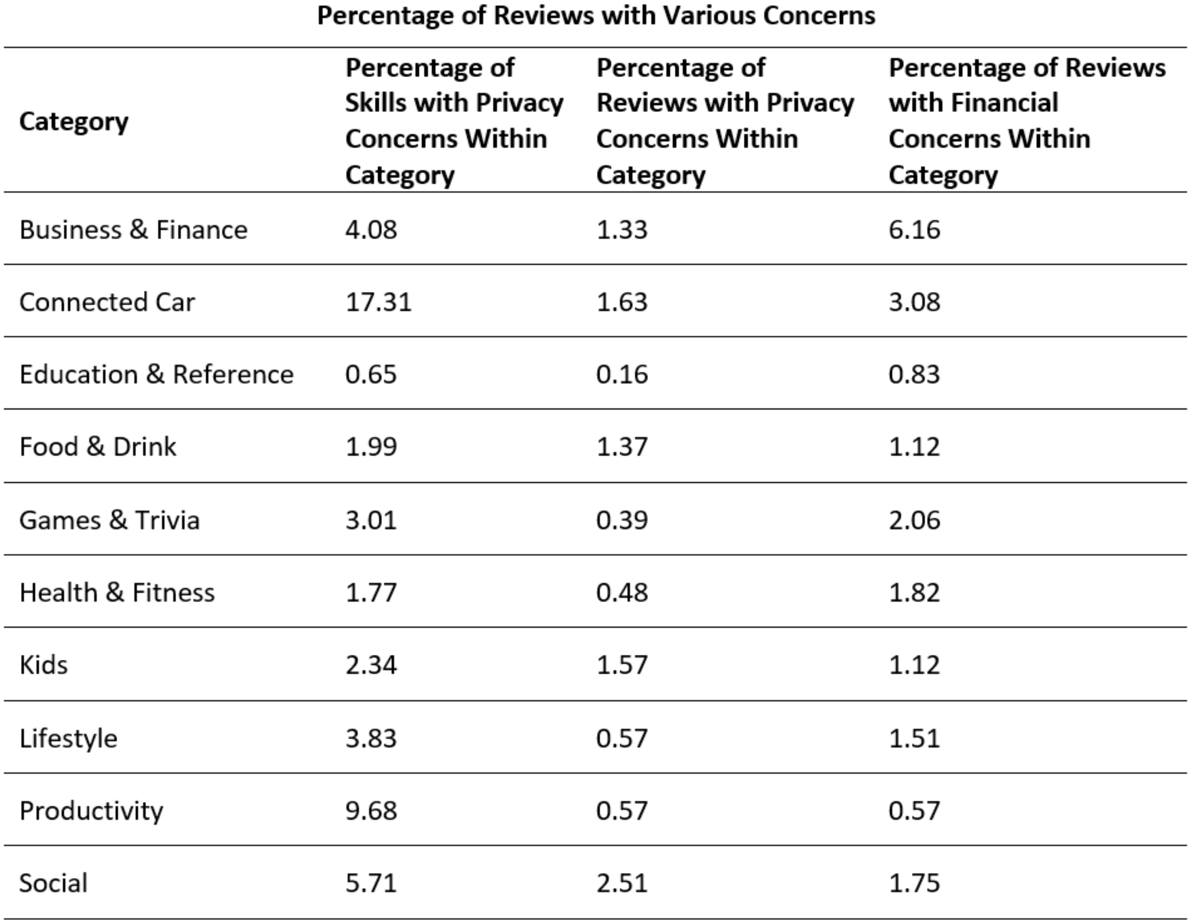


Figure 6: Number of skills with privacy complaints and their percentages

Frequency of Concerns

The presence of privacy concerns in reviews and skills are fairly well distributed across all categories given the proportion of total reviews/skills each category occupies. However, the frequency of these reviews appear to be fairly low in this dataset, as shown in the second column of Figure 7. The percentages of reviews with privacy concerns are all below 3%, with the highest being 2.51% in the "Social" category. This suggests that privacy is not necessarily a frequent problem, but a widespread one.



Percentage of Reviews with Various Concerns

Category	Percentage of Skills with Privacy Concerns Within Category	Percentage of Reviews with Privacy Concerns Within Category	Percentage of Reviews with Financial Concerns Within Category
Business & Finance	4.08	1.33	6.16
Connected Car	17.31	1.63	3.08
Education & Reference	0.65	0.16	0.83
Food & Drink	1.99	1.37	1.12
Games & Trivia	3.01	0.39	2.06
Health & Fitness	1.77	0.48	1.82
Kids	2.34	1.57	1.12
Lifestyle	3.83	0.57	1.51
Productivity	9.68	0.57	0.57
Social	5.71	2.51	1.75

Figure 7: Frequency of various concerns by category

From Figure 7, comparing the percentage of reviews with privacy concerns and the percentage of reviews with financial concerns reveals a notable trend. Nearly all categories had a higher percentage of reviews with financial concerns except for three categories: “Food & Drink”, “Kids” and “Social”. These results suggest that Alexa skill reviewers are particularly concerned about privacy while using “Social” skills, and “Kids” skills. Upon further manual examination of the dataset, privacy concerns in the “Food & Drink” category were primarily related to scams that attempted to collect personal information from the user. Privacy concerns in the “Social” category were generally concerns regarding sharing information with the social network or with the skill without consent. Reviewers with privacy complaints in “Kids” skills expressed that these skills attempted to gather and sell information about the children using the skill.

Conclusion

Some of the most common issues mentioned by Alexa skill reviewers were inappropriate content, financial concerns about the skill, and privacy and security concerns. Reviews regarding inappropriate content were fairly rare, with the most mentions of inappropriate content being in the "Games & Trivia" category of skills. Financial complaints were found to be much more common across all categories. "Games & Trivia" dominated negative finance-related reviews, with 38% of these reviews being in the "Games & Trivia" category. Privacy and security concerns were present across all categories of Alexa skills examined in this data set, though less frequently than financial concerns. The "Kids" category was particularly interesting as it appears to have a handful of skills that are particularly unsafe with many negative reviews. Privacy concerns are also more common in "Social", "Kids", and "Food & Drink" categories than financial issues. There are many possible solutions to resolve some of these issues.

Potential Solutions

These results suggest that the "Kids" category may need more content moderation to ensure that Alexa skills do not attempt to ask children personal information. The "Social" category may need to require skills to implement some form of authentication before sharing information across social networks, and give users an option to opt out of sharing certain personal information. Transparent privacy policies are another potential remedy to these issues present in "Social" skills. On the other hand, "Games & Trivia" need better content moderation to check for inappropriate content, particularly in games that are supposedly safe for children.

Similar to the “Social” skills, these skills in “Games & Trivia” would also benefit by implementing an authentication mechanism before users accidentally pay fees to use the skill.

Future Work

With the limitations of the research described before, there are many ways to expand both the scope and accuracy of this research. A natural way to expand the scope is acquire a larger data set to include more skills or more categories. There are some categories that were not included in this data set such as Weather, which we may expect to see complaints about sharing location data. With more data, it would also be possible to identify and sort the different types of privacy complaints, and to compare their frequencies. For example, are there privacy concerns about conversations being recorded? Or are there more concerns about stealing credit card information? What about names, birthdays, and location information? To improve the accuracy of the results, a more sophisticated natural language processing algorithm could be devised instead of simply matching a set of words. This would reduce the amount of false positives and false negatives.

Future research could also identify other sources of complaints, such as skills not working correctly, and compare the frequency of such complaints to the frequency of security concerns. This could provide further insight into how frequent these complaints are in comparison to other issues. Expanding on this idea, it would also be interesting to examine these privacy complaints in relation to the Alexa Privacy Policy and the privacy policy for each skill. Are these skills violating the privacy policy? Are they not? Does Amazon Alexa need to implement or improve a general privacy policy that applies to all skills? Drawing such conclusions from research could help strengthen privacy in all Alexa skills, and in particular, better protect vulnerable groups such as children from giving personal information.

References

Pal, D., Arpnikanondt, C., Razzaque, M. A., & Funilkul, S. (2020). To Trust or Not-Trust: Privacy Issues With Voice Assistants. IT Professional, 22(5), 46-53.

Pridmore, J., & Mols, A. (2020). Personal choices and situated data: Privacy negotiations and the acceptance of household Intelligent Personal Assistants. Big Data & Society, 7(1), 2053951719891748. Web of Science.

Sacks, E. (2018, May 26). Alexa privacy fail highlights risks of smart speakers. NBC News. https://www.nbcnews.com/tech/innovation/alexa-privacy-fail-highlights-risks-smart-speakers-n877671