Alzate, M., Arce Urriza, M. and Cortiñas, M. (2023), "Voice-activated personal assistants and privacy concerns: a Twitter analysis", Journal of Research in Interactive Marketing, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JRIM-02-2023-0050

Voice Activated Personal Assistants and Privacy Concerns: A Twitter Analysis

1. Introduction

Voice Activated Personal Assistants (VAPAs) such as Siri, Alexa, and Google Assistant have become integral to our daily lives. These tools, which are often built into smartphones or other smart devices, serve various functions ranging from web searches to device control. VAPAs present two important characteristics. First, they are a type of Artificial Intelligence (AI)- Internet of things (IoT) device that enable an easy and fast two-way interactive communication between users and providers of services, what facilitates a more interactive, tailored and personalized relationship between them (Wang, 2021; Wang, 2023). Second, VAPAs are one of the few products in the market to use the human voice as the input to control the functionalities of a device (Chen et al., 2022). Using voice as the main human-machine interface allows a more humanized relationship. Han and Yang (2018) deeply study the user acceptance of VAPAS evaluating the social characteristics associated to them. They find that considering the assistant as a friend, i.e., establishing an interactive and socially enjoyable relationship with it, increases satisfaction and the intention to continue in the use of VAPAs. In fact, the more the users interact with VAPAs, the greater they engage with them (Chen et al., 2022). Besides, as Jang et al. (2022) recall, the anthropomorphism of VAPAs does not only provide interactivity with the customer but also helps to customize the search process, what can facilitate voice shopping, increasing customer satisfaction and opinions towards smart speakers.

One of the most important and recent concerns regarding the use of VAPAs is the increasing preoccupation with the use of personal data and privacy protection policies (Cao & Wang, 2022; Massara, 2021). These concerns are further exacerbated by notable incidents of misuse of personal data by major VAPA providers.[1] The magnitude of this privacy preoccupation and its dimensions is still unknown and misunderstood, as reflected in the calls to study the potential dark side of interactive and personalized technologies, which VAPAs are a type of (e.g., Wang, 2021). The results of Han and Yang (2018) also highlight this preoccupation, given that security and privacy risks negatively affect the parasocial relationship with VAPAs.

This paper aims to shed light on consumers' privacy concerns regarding VAPAs, drawing from social exchange (Ashworth & Free, 2006; Malhotra et al., 2004) and Privacy Calculus theories (Jiang et al., 2013). Both theories rely on the fact that, when setting boundaries of information disclosure with VAPAs, individuals have to determine whether revealing private information is worthwhile considering the potential benefits and associated risks. This research focuses on three research questions:

RQ1) Does press coverage of privacy impact the volume and sentiment of conversations about VAPAs?

RQ2) Are the sentiments of conversations about privacy issues with VAPAs generally more negative than those of general conversations about VAPAs?

RQ3) What specific privacy-related issues are most frequently discussed in conversations about VAPAs and how do these issues impact the volume and sentiment of these conversations over time?

---

[1] See for example https://www.theguardian.com/technology/2019/aug/29/apple-apologises-listen-siri-recordings; https://edition.cnn.com/2019/04/11/tech/amazon-alexa-listening/index.html , https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms/

Previous studies have demonstrated that perceived risks and privacy concerns play a crucial role in consumer decision-making and attitudes towards voice assistants (Kowalczuk, 2018; Taylor et al., 2009; Zhou, 2013). However, not much research deepens in the study of what specific privacy concerns are associated with VAPAs. Moreover, literature in communication has revealed not only that consumers´ exposure to news impacts their mood and the type of content they publish in social media (Hasell, 2021) but also that consumers might be even more influenced by negative information (Park, 2015; Zhu et al., 2020). In this research, we analyze if this also happens when consumers are exposed to privacy-related news about VAPAs. Overall, this study enhances our understanding of the evolving landscape of privacy concerns in the context of emerging technologies like VAPAs, with implications for both theory and practice.

Our research involved analyzing two years of Twitter discussions on Amazon Alexa, Google Assistant, and Apple Siri, encompassing 441,427 tweets. We conducted text mining analysis to reveal the sentiment of the tweets over the period. We also monitored privacy-centric news on these VAPAs during this period.

2. Theoretical Background

2.1.Information privacy and VAPAs

Information privacy, defined as the control that individuals have over their personal data (Westin, 1967; Campbell, 1997), has been primarily studied in the context of direct marketing rather than online channels (Cao & Wang, 2022). However, the interactive nature of the Internet increases privacy risks (Malhotra et al., 2004; Sheehan & Hoy, 2000). With the proliferation of e-commerce, companies have accumulated extensive consumer data, intensifying concerns about online privacy (Ashworth & Free, 2006; Malhotra et al., 2004; Phelps et al., 2000).

The advent of AI, particularly Voice-Activated Personal Assistants (VAPAs), has further complicated privacy issues. These devices rely on extensive personal and contextual data to function (Cao & Wang, 2022). Massara (2021) noted that two-thirds of Europeans are concerned about the lack of information control. Because VAPAs collect sensitive data, including user preferences and demographic information, they pose significant risks of data leakage and privacy abuse. Thus, information privacy remains a critical concern in AI and e-commerce.

When interacting with humans and artificial conversational agents, individuals choose to disclose or retain private information to control accessibility (Petronio, 1991). Two main perspectives guide these choices: economic rationality and social exchange, both of which acknowledge the inherent risk of information loss (Cao & Wang, 2022, Malhotra et al., 2004).

From an economic rationality perspective, the Privacy Calculus theory posits that disclosure decisions are based on cost-benefit analyses. If perceived benefits, such as VAPAs' personalized services, exceed potential risks, individuals are more likely to disclose information (Jiang et al., 2013; Kehr et al., 2015; Li et al., 2021).

Conversely, the Social Contract (SC) theory, emanating from the social exchange perspective, identifies three key factors affecting consumer privacy concerns: collection, control, and awareness (Malhotra et al., 2004; Donaldson & Dunfee, 1994). 'Collection' entails balancing the amount of personal information shared against the received benefits, emphasizing equitable and respectful data gathering. 'Control' underscores an individual's right to dictate how their personal information is used (Ashworth & Free, 2006; Malhotra et al., 2004). 'Awareness' advocates transparency and understanding of how personal data are managed (Shapiro et al., 1994). As explained by Cao and Wang (2022), users establish information-sharing relationships when dealing with interactions with VAPAs. Considering the concept of privacy as a "boundary," users should decide and form the rules of disclosing or not privacy information to VAPAs.

During their interactions with VAPAs, users establish information-sharing relationships. The decision to disclose information to the VAPAs should consider these multifaceted privacy theories. By understanding privacy as a "boundary," users form rules governing the information they share, weighing the benefits, costs, and autonomy of these transactions (Cao & Wang, 2022). This synthesis of theories aids in navigating the complex landscape of information privacy in the era of advanced conversational technology.

Henkens et al. (2020) found that smarter products improve user experience, but also increase perceived intrusion. Ameen et al. (2021) cited privacy as part of the "perceived sacrifice" involved in using AI technologies, encompassing loss of control and effort. Lucia-Palacios and Pérez-López (2023) suggested that, as smart products become more autonomous, perceived risks or losses increase, affecting the product's perceived value. Lee et al. (2020) placed privacy within the broader framework of perceived security and satisfaction. In this same line, Hsieh and Lee (2021) found that trust on the smart speaker lead to a positive attitude toward using them.

Lau, Zimmerman, and Schaub (2018) noted that the adoption of VAPAs varies according to individual perceptions of privacy and utility. Non-users often avoid VAPAs owing to insufficient perceived benefits or significant privacy concerns. Kudina and Coeckelbergh (2021) further explored such fears, finding that, despite general mistrust, many users continue to use these devices, believing that the benefits outweigh the risks. Both studies call for more comprehensive research to better understand consumers' complex privacy concerns related to VAPAs.

The evaluation of privacy risks also differs among VAPAs. CommonSense (2019) rated Siri higher on transparent privacy policies than Google Assistant and Amazon Alexa. These perceptions may be related to the awareness factor in the social exchange theory of privacy (Malhotra et al., 2004), indicating that transparency influences consumer trust.

2.2.VAPAs and social buzz

The Internet, particularly social media, has become a significant platform for shaping public opinion on products and services, including VAPAs. Researchers have used platforms like Twitter to explore diverse aspects, such as consumer engagement, firm-customer relationships, and public sentiment (Abbasi et al., 2019; Chung et al., 2020; Novak & Vilceanu, 2019). In particular, Twitter has been highlighted as a crucial platform through which companies can monitor customer opinions and engage in meaningful dialogue (Eriksson, 2018; Xu & Wu, 2015). Besides, the impact of news shared on such platforms can be long lasting and varies across different customer segments and companies (Scurlock et al., 2020; Septianto, 2020).

The literature shows that the frequency of news articles on a topic such as privacy concerns influences the volume of related social media conversations (Arifin & Lennerfors, 2022; Hasell, 2021). Moreover, the emotional intensity of the language used in these posts affected their distribution and impact. While emotional language has generally garnered more attention, its role in information quality and diffusion remains debatable (Huffaker, 2010; Zhang and Peng, 2015; She et al., 2022).

Particularly noteworthy is the phenomenon of "negativity bias," where negative content tends to have a stronger impact than positive content, both in traditional and social media (Park, 2015; Zhu et al., 2020; Garz, 2014; Soroka et al., 2018). However, She et al. (2022) found that, although negative emotional content leaves a deeper impression, both positive and negative emotions attract readers, albeit the effect is stronger in the case of negative emotions. Conversely, some studies suggest that positive content could have a greater impact on social media, although its diffusion might not generate similar emotional responses (Goldenberg & Gross, 2020; X. Wang & Lee, 2020).

2.3.Research questions

In light of the literature review, and to address these gaps, three research questions were articulated.

RQ1) Does press coverage of privacy impact the volume and sentiment of conversations about VAPAs?

Some studies (Arifin & Lennerfors, 2022; Hasell, 2021) have suggested that a higher volume of information about VAPAs, irrespective of their positive or negative nature, generates more social buzz. Additionally, a "negativity bias" (Park, 2015; Zhu et al., 2020) has been discussed, indicating that negative news might disproportionately affect public sentiment. Similarly, research relying on the privacy-satisfaction model supports the notion that greater feelings of privacy lead to increased user satisfaction (Jang et al., 2022; Lee et al., 2020). Applied to our context, this suggests that the preponderance of negative news about VAPA privacy can deteriorate public sentiment, which is equated with user satisfaction. However, there is a lack of research on the relationship between media communication and privacy, which makes delving into this research interesting and relevant.

RQ2) Are the sentiments of conversations about privacy issues with VAPAs generally more negative than those of general conversations about VAPAs?

Existing literature posits that perceived risks, including privacy concerns, negatively influence the perceived value of a product (Park et al., 2018, Taylor et al., 2009; Zhou, 2013). Research focused on VAPA adoption has shown that these perceived risks reduce social attachment to devices (Han & Yang, 2018) and diminish user intentions (Kowalczuk, 2018). Lau et al. (2018) highlighted that concerns over the use of private data largely deter the adoption of VAPA. Moreover, the skepticism often stems from unclear privacy policies of companies and a general lack of consumer awareness (Hui et al., 2007; Malhotra et al., 2004). Given these findings, it is reasonable to anticipate that discussions featuring privacy concerns regarding VAPAs will generate more negative sentiments, reflecting the associated perceived risks.

RQ3) What specific privacy-related issues are most frequently discussed in conversations about VAPAs and how do these issues impact the volume and sentiment of these conversations over time?

Most previous research have focused on analyzing privacy through scales in questionnaires (Han & Yang, 2018; Kowalczuk, 2018; K. Park et al., 2018), which might restrict consumers to raise their real feelings about VAPAs privacy. Another stream or research, such as Lau et al. (2018) and Kudina and Coeckelbergh (2021), primarily adopts an exploratory approach through interviews, providing only initial insights into specific consumer concerns about privacy. However, there's limited evidence from studies on privacy concerns related to VAPAs. More research is needed to understand the specific issues consumers discuss and how it affects their sentiments and conversation volumes.

3. Methodology

3.1. Data collection

We used data from Twitter to analyze privacy concerns regarding VAPAs. There were several reasons for this choice. Twitter is used globally and is particularly relevant to tech-savvy demographics that use VAPAs. The platform allows the immediate expression of opinions, captures authentic consumer views, and accommodates both VAPA users and non-users.

We collected tweets mentioning Amazon Alexa, Google Assistant, and Apple Siri using the Twitter API and the ‘rtweet’ package in R from July 1, 2019, to June 30, 2021. To navigate API restrictions, we performed weekly searches focusing on tweets containing key terms related to the three VAPAs, excluding retweets. Our initial dataset comprised 638,612 tweets. After data cleaning, which involved removing tweets with fewer than four words and duplicate texts, 441,427 tweets remained for analysis.

Additionally, we monitored the privacy-related press coverage of the three VAPAs during the study period. We conducted Google searches using relevant keywords like "privacy," "google," "siri," and "alexa" to compile a dataset of news articles published in the same period. They were categorized as either positive or negative based on their content. News sources ranged from mainstream media such as The Washington Post to specialized tech forums and even the brands themselves. Data were aggregated at both the daily and weekly levels, providing a comprehensive overview of public sentiment and buzz. Our methodology aligns with that of Arifin and Lennerfords (2022), who examined the ethical aspects of VAPAs by analyzing media discourse. The authors collected articles using similar keywords and followed a parallel procedure for data collection and analysis.

3.2. Composition of the database

Figure 1 illustrates the distribution of tweets mentioning "privacy" over the study period. The proportion of privacy-related tweets was relatively low compared with that of the entire dataset (2.1%). However, this should not undermine their significance, as studies such as Maccario and Naldi (2022) and Manikonda et al. (2018) emphasize that privacy is a crucial factor in consumers' decisions to adopt VAPAs. Conversations about privacy were more prevalent at the beginning of the study period (roughly between weeks 0 and 18). Among the VAPAs, Apple Siri had the highest proportion of privacy-related tweets, followed by Amazon Alexa, with Google Assistant registering the fewest. To interpret these time-based variations, our first research question aimed to assess the effect of privacy-focused press coverage on weekly tweet volume and sentiment for each VAPA.

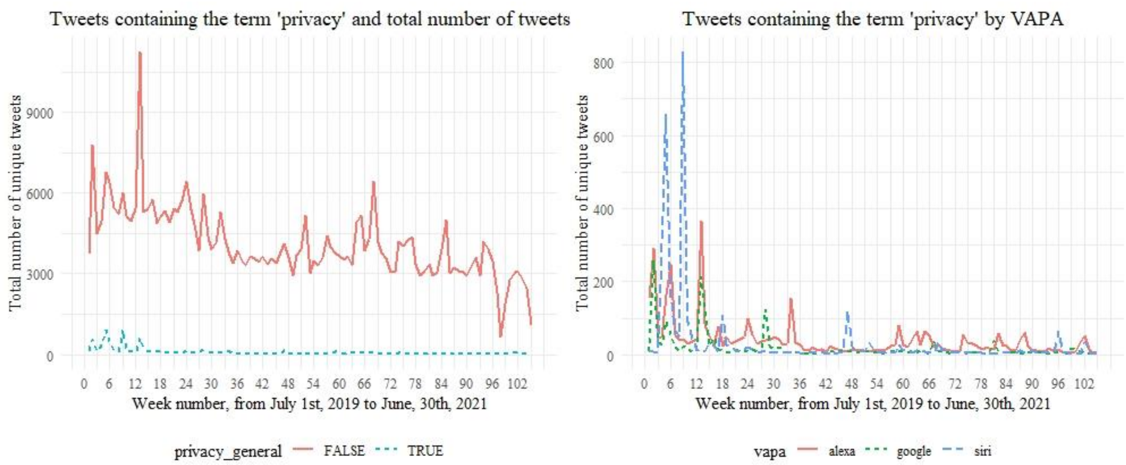

Figure 1. Number of tweets containing the term "privacy" over the total amount of tweets (in general) and number of tweets containing the term "privacy" by VAPA

3.3. Textual analysis of tweets

Before processing the tweets, we cleaned the texts to avoid bias in the detection of sentiments and issues. That process of cleaning involves removing URLs, mentions to other users using the symbol "@", hashtags, punctuation, and other types of symbols. We adopted a lexicon-based approach for sentiment analysis, Linguistic Inquiry and Word Count (LIWC). The LIWC was initially developed by Pennebaker et al. (2001) and, since its introduction, has been validated in many studies in different psychological and marketing domains that analyze content, such as instant messaging and online blogs (Ireland et al., 2011; Ludwig et al., 2013). As stated by Alzate et al. (2022), the linguistic indicator scores for each LIWC variable were calculated as the percentage of words matching the predefined dictionary.

To estimate the sentiment of each tweet, three measures from LIWC 2022 dictionary were used: "affect," "tone," and "risk" (Boyd et al., 2022). LIWC22 is proprietary software that has not released the exact way these variables are calculated, but general descriptions can be found in Boyd et al. (2022). Of the three measures used, "affect" "*reflects sentiment reference to emotions and includes words related to positive and negative emotion (e.g., happy, joy, sad, angry) and also words related to those emotions (e.g., birthday, beautiful, kill, funeral)…, while "Emotional Tone" words are restricted to true emotion labels and words that strongly imply emotions*" (Boyd et al, 2022, page 18). "Risk" includes words related to danger and uncertainty. It includes words that signal potential harm or loss, such as "danger," "doubt," "risk," and "guess." Finally, We incorporated VADER's "compound" score (Hutto & Gilbert, 2014), which which calculates the sentiment of words and normalizes it between -1 (very negative) and +1 (very positive).

Besides using LIWC22 and VADER for sentiment analysis, we employed a natural language processing (NLP) tool to identify main keywords in private conversations. Keyword extraction, a type of text mining analysis tool, aims to automatically pinpoint key concepts in texts.

R software was used for this text mining task using packages such as dplyr (Wickham & Francois, 2016) and tidyr (Wickham, 2016).

4. Results

This section is organized to answer each of our research questions.

4.1. RQ1) Does press coverage of privacy impact the volume and sentiment of conversations about VAPAs?

We used two different models: Model 1 analyzed the impact of press coverage on the median sentiment of conversations per week and Model 2 analyzed the impact of press coverage on the number of tweets per week.

Model 1 is specified as follows:

$$Sentiment_{vw} = Google + Siri + Pos\_News_{vw} + Neg\_News_{vw} + Sentiment_{vw-1} + Year_{w}$$

- $Sentiment_{vw}$ is the median of the sentiments of all tweets for each VAPA (v) aggregated to a weekly level (w).

- *Google* and *Siri* are dummy variables that indicate VAPA, the reference level is *Alexa*.

- Pos_News_{vw} is the number of positive news released for each VAPA (v) per week (w).

- Neg_News_{vw} is the number of negative news released for each VAPA (v) per week (w).

Two additional control variables were also included. First, $Sentiment_{vw-1}$ is the median sentiment of all tweets for each VAPA (*v*) in the previous week (*w-1*). This variable was included to control the lagged effects of past negative feelings. Finally, $Year_w$ controls for the year in a specific week (*w*).

To compute the sentiment of each tweet, as described previously, we used four different measures of sentiment: affect, tone, and risk from the same LIWC22 dictionary and the compound measure from VADER. The results are similar for all four dependent variables. Here, we present the results for the variable Affect, from LIWC22.[2]

Model 2 analyzed the impact of press coverage on the number of tweets specified as follows:

$$N\_Tweets_{vw} = VAPA_v + Pos\_News_{vw} + Neg\_News_{vw} + N\_Tweetst_{vw-1} + Year_w$$

In this case, the independent variables were dummy variables for assistants, number of positive and negative news items, lag number of tweets, and year.

More specifically

- N_Tweets_{vw} is the total number of tweets for each VAPA (*v*) aggregated weekly (*w*).

- *Google* and *Siri* are dummy variables that indicate VAPA, the reference level is *Alexa.*

- Pos_News_{vw} is the number of positive news released for each VAPA (*v*) per week (*w*).

[2] The rest of the models are available from authors upon request.

- Neg_News_{vw} is the number of negative news released for each VAPA (*v*) per week (*w*).

- N_Tweets_{vw-1} is the lag of the number of tweets for each VAPA (*v*) of the previous week (*w-1*).

- $Year_w$ the year of that specific week (*w*).

Table 1 about here

Table 1 shows the results of the multivariate linear regression for Models 1 and 2. The number of observations for both models was 306 (102 weeks, three VAPAS). The adjusted R2 values for both models were reasonable, particularly for Model 2 in which this specification accounted for approximately 60% of the variance in the number of tweets.

In Model 1, it can be observed that the number of negative news about privacy has an impact on the sentiment of the tweets (β = -0.30, sig = 0.003). Therefore, the higher the volume of negative press coverage on privacy in a week, the more negative the sentiment of the tweets. On the contrary, one can notice that positive news has not significant impact on the sentiment (β = -0.16, sig = 0.32). Thus, negative privacy news causes consumers to change their discourse on VAPA, but positive news does not have that power. We also observed a negative effect of Google Assistant and Siri in comparison with the omitted dummy (Alexa), indicating that both assistants generally have a more negative conversation than that about Alexa. We found no evidence of a lag effect of past sentiment in the current week. Finally, we found significant and positive effects for the dummies for 2020 and 2021, indicating that the most negative conversations occurred in 2019.

Model 2 shows that both the number of negative and positive news about privacy have a positive effect on the number of tweets ($\beta = 260$, sig = 0.011 and $\beta = 213$, sig < 0.001). Thus, the higher the number of positive or negative privacy-related news items about VAPAs in a week, the higher the number of tweets written by consumers about VAPA. All news increases the volume of conversations about VAPAs. We also observed a negative effect of Google Assistant and Siri in comparison with the omitted dummy (Alexa), indicating that both assistants had fewer tweets than Alexa. In Model 2, as in Model 1, we found no evidence of the effect of the lag-dependent variable. As far as the year effect is concerned, we can observe that the conversation in 2019, when most VAPA scandals were exposed, was also more abundant than in 2020 and 2021. Thus, the effect of those "scandal" news stories diminished over time.

4.2. RQ2. Are the sentiments of conversations about privacy issues with VAPAs generally more negative than those of general conversations about VAPAs?

Our theory suggested that tweets with the words "privacy" or "private" would be more negative. To test this, we used four sentiment measures from two dictionaries to ensure unbiased results based on how "sentiment" is defined.

Table 2 about here

We estimated four ANOVA models, where the dependent variable is the sentiment of the tweet and the independent variable is a dummy variable for tweets containing privacy words. We compared the sentiments of privacy-related tweets (9,341 tweets) to the sentiments of the remaining tweets (432,186 tweets). The results of the ANOVA models with different approaches to the dependent variable of sentiment are shown in Table 2.

The analysis reveals that, regardless of how we measure sentiment, the average sentiment of privacy-related tweets is significantly more negative than that of the rest of the tweets. We observed that the three different measures of general sentiment (affect, tone from LIWC22, and compound measure from VADER) were lower in privacy tweets than in the other tweets, indicating a more negative conversation. In the specific case of the measure of risk, the value was more than four times higher for tweets mentioning privacy than for the remaining tweets (0.69 vs 0.16), indicating that the conversation in these tweets involves more risk-related concepts.

4.3. RQ3. What specific privacy-related issues are most frequently discussed in conversations about VAPAs, and how do these issues impact the volume and sentiment of these conversations over time?

In our last research question, we intended to further investigate the specific issues associated with privacy to determine the differential effects of each. For this analysis, we focused on a subset of privacy-related tweets (7,994 tweets). We extracted the most frequent words from this subset of tweets and selected the tenth most frequent terms: *Record, Voice, Data, Listen, User, Conversation, Concern, Contractor, Home,* and *Device*. Appendix A presents a complete table of keywords.

After identifying the keywords of interest, we searched for them in the dataset. The following two models were established:

The first model is specified as follows:

$$Sentiment_t = Google \quad + Siri + mention\_term1_t + mention\_term2_t + \cdots + mention\_term10_t + Year_t \qquad \text{(Model 1)}$$

The dependent variable is the tweet t sentiment, measured by the affect value provided by LIWC, and the independent variables are the dummy variables for the assistants, the dummy variables for the mentions of each of the ten terms of interest in a tweet t, and the year in which the tweet t was published.

The second model is specified as follows:

$$N\_tweets_{dv} = Google \quad + Siri + mention\_term1_{dv} + mention\_term2_{dv} + \cdots + mention\_term10_{dv} + Ntweets_{vd-1} + Year_{t} \qquad \text{(Model 2)}$$

The dependent variable is the number of privacy-related tweets published each day d for each assistant v, and the independent variables are the dummy variables for the assistants, the number of tweets mentioning each of the ten terms of interest each day d for each assistant v, the lag value of the number of privacy-related tweets in the previous day for each assistant v, and the year to which day d belongs.

The two models, shown in Table 3, were estimated using linear regression to determine the keywords that had a greater impact on conversation sentiment (Model 1) and volume (Model 2). The number of observations was 7994 for the first model (total number of tweets mentioning privacy in the database) and 1509 for the second model (number of daily aggregations of tweets mentioning privacy for each of the VAPAS).[3]

Model 1 in Table 3 shows the results of the regression model for tweet sentiment, where we analyzed the effect of the number of tweets mentioning each keyword on sentiment. Model 2 in Table 3 shows the results of the regression model with the number of daily tweets as the dependent variable.

Table 3 about here

---

[3] 700 days of tweets in our database and three assistants would have made 2,100 observations. However, not every day in the database there are tweets mentioning each of the three assistants. There are some days without tweets mentioning privacy, other days where tweets mention privacy for only one or two VAPAs and other days where we can find privacy-related tweets for the three VAPAs.

From the results of Model 1, several terms—namely, "concern," "record," "listen," "conversation," "contractor," "data," and "voice"—have been identified as significant predictors of sentiment. Interestingly, these terms carry a negative coefficient, implying that their presence in a tweet generally correlates with a decrease in the sentiment score. In other words, tweets containing these terms, which are all related to privacy issues, tended to demonstrate more negative sentiments. Additionally, the term "home" was associated with a positive impact on tweet sentiment, suggesting tweets that mention "home" tend to reflect a more positive sentiment.

As for the variable representing the type of device (Alexa, Siri, or Google), no notable differences were observed in the average sentiment scores. This suggests that the type of device under discussion did not significantly alter the overall sentiment of the tweet.

Looking at the year-wise comparison, the sentiments in tweets from 2020 and 2021 appear to be more negative than those in tweets from 2019. This observation reflects an increasing trend of negative sentiments in the domain of our study over the years for tweets talking about privacy.

Model 2 in Table 3 presents the findings of the regression model in which the dependent variable is the daily volume of tweets throughout the study period. As per these results, the mention of several terms—"device," "record," "data," "user," "concern," "conversation," "voice," and "listen"—significantly influences the daily number of tweets. These terms demonstrate a positive coefficient, suggesting that their mentions tend to coincide with an increased daily volume of tweets, listed in decreasing order of impact.

By contrast, the term "contractor" shows a negative correlation with the daily tweet count. This implies that on days when "contractor" is more frequently mentioned, there is a tendency for fewer tweets to be posted.

Furthermore, neither the type of device (Alexa, Siri, or Google) nor the reference year seemed to impact the daily number of tweets, suggesting a consistent level of discussion regardless of these factors.

When comparing Model 2 to Model 1, terms affect sentiment and tweet volume differently. For instance, the term "concern" decreases sentiment but increases the volume of tweets, suggesting that although its mention tends to generate more discussions, those conversations are usually negatively charged. The term "contractor" has a similar pattern, reducing both sentiment and tweet volume, indicating less frequent but generally more negative discourse when this term appears. The term "home," which improves sentiment, does not impact tweet volume, suggesting that positive discussions involving "home" are not necessarily more frequent**.**

Device type, whether Alexa, Siri, or Google, doesn't affect sentiment or tweet volume in either model. However, there is a contrasting effect between the two models with respect to the reference year. In the sentiment analysis (Model 1), the year had a significant impact, with more negative sentiments observed in 2020 and 2021 than in 2019. However, in the tweet volume analysis (Model 2), the reference year did not significantly affect the daily number of tweets, suggesting a stable level of discussion over time.

5. Discussion

5.1 General discussion

This study analyzes the sentiments associated with privacy-related conversations, the effect of VAPAs positive and negative news on those conversations second, and the specific privacy risks linked to VAPAs as perceived by consumers. To better link the main findings of this study to previous literature and to managerial implications, we developed Table 4.

Table 4 about here

5.2 Theoretical Contribution

This paper builds on previous research about online privacy and voice assistants (VAPAs) to understand consumer privacy concerns. Using social exchange (Ashworth & Free, 2006; Malhotra et al., 2004) and privacy calculus theories (Jiang et al., 2013), we explore the specific privacy risks associated with VAPAs. While past studies highlight the importance of privacy concerns in shaping user attitudes towards voice assistants, there's a gap in understanding specific concerns. Our study delves deeper into these risks, examining their impact on consumer emotions and the amount of related social discourse

Moreover, communication studies show that consumers are swayed by news content, especially negative ones (C. S. Park, 2015; Zhu et al., 2020). Our research builds on this by investigating if this bias continues when consumers encounter varied press coverage about privacy issues related to VAPAs

This study employs a unique method by analyzing privacy concerns through social buzz, a technique effective for large consumer data. As Kozinets et al. (2010) highlighted, consumers tend to share genuine perceptions and behaviors about products and brands in such settings.

5.3 Limitations and future research directions

Our study had some limitations. It is based on Twitter data and excludes other platforms and offline opinions. The study also limits the scope to tweets mentioning "privacy," potentially overlooking related terms like "security" or "personal data.". Furthermore, we focused on the sentiments of the general Twitter population, rather than distinguishing between VAPA users and non-users or varying levels of expertise.

In future research, multiple methodologies such as surveys and online reviews could offer a fuller understanding. Additionally, this study could expand its scope to include various terminologies, compare sentiments across cultures, and consider the gravity of news in shaping public opinion.

Despite these limitations, our study opens new avenues for understanding the dynamics of press coverage, consumer sentiment, and the volume of social discourse on VAPAs. In doing so, it offers both scholarly and practical insights into the evolving nature of consumers' privacy concerns.

6. References

Abbasi, A., Li, J., Adjeroh, D., Abate, M., & Zheng, W. (2019). Don't Mention It? Analyzing User-Generated Content Signals for Early Adverse Event Warnings. *Information Systems Research*, *30*(3), 1007–1028. https://doi.org/10.1287/isre.2019.0847

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2022). Mining the text of online consumer reviews to analyze brand image and brand positioning. *Journal of Retailing and Consumer Services*, *67*, 102989. https://doi.org/10.1016/j.jretconser.2022.102989

Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, *114*(June 2020), 106548. https://doi.org/10.1016/j.chb.2020.106548

Arifin, A. A., & Lennerfors, T. T. (2022). Ethical aspects of voice assistants: A critical discourse analysis of Indonesian media texts. *Journal of Information, Communication and Ethics in Society*, *20*(1), 18–36. https://doi.org/10.1108/JICES-12-2020-0118

Ashworth, L., & Free, C. (2006). Marketing Dataveillance and Digital Privacy: Using Theories of Justice to Understand Consumers' Online Privacy Concerns. *Journal of Business Ethics*, *67*(2), 107–123. https://doi.org/10.1007/s10551-006-9007-7

Boyd, R., Ashokkumar, A., Seraj, S., & Pennebaker, J. (2022). *The Development and Psychometric Properties of LIWC-22*. https://doi.org/10.13140/RG.2.2.23890.43205

Campbell, A. J. (1997). Relationship marketing in consumer markets: A comparison of managerial and consumer attitudes about information privacy. *Journal of Direct Marketing*, *11*(3), 44–57. https://doi.org/10.1002/(SICI)1522-7138(199722)

Cao, G., & Wang, P. (2022). Revealing or concealing: Privacy information disclosure in intelligent voice assistant usage- a configurational approach. *Industrial Management & Data Systems*, *122*(5), 1215–1245. https://doi.org/10.1108/IMDS-08-2021-0485

Chen, Y. H., Keng, C.-J., & Chen, Y.-L. (2022). How interaction experience enhances customer engagement in smart speaker devices? The moderation of gendered voice and product smartness. *Journal of Research in Interactive Marketing*, *16*(3), 403–419. https://doi.org/10.1108/JRIM-03-2021-0064

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2020). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–

595. https://doi.org/10.1016/j.jbusres.2018.10.004

Combs, M., Hazelwood, C., & Joyce, R. (2022). Are you listening? – An observational wake word privacy study. *Organizational Cybersecurity Journal: Practice, Process and People*, *2*(2), 113–123. https://doi.org/10.1108/OCJ-12-2021-0036

Commonsense. (2109). *Compare the Privacy Practices of the Most Popular Smart Speakers with Virtual Assistants*. Common Sense Education. https://www.commonsense.org/education/articles/compare-the-privacy-practices-of-the-most-popular-smart-speakers-with-virtual-assistants

Donaldson, T., & Dunfee, T. W. (1994). Toward a Unified Conception of Business Ethics: Integrative Social Contracts Theory. *Academy of Management Review*, *19*(2), 252–284.

Eriksson, M. (2018). Lessons for Crisis Communication on Social Media: A Systematic Review of What Research Tells the Practice. *International Journal of Strategic Communication*, *12*(5), 526–551. https://doi.org/10.1080/1553118X.2018.1510405

Garz, M. (2014). Good news and bad news: Evidence of media bias in unemployment reports. *Public Choice*, *161*(3), 499–515. https://doi.org/10.1007/s11127-014-0182-2

Goldenberg, A., & Gross, J. J. (2020). Digital Emotion Contagion. *Trends in Cognitive Sciences*, *24*(4), 316–328. https://doi.org/10.1016/j.tics.2020.01.009

Han, S., & Yang, H. (2018). Understanding adoption of intelligent personal assistants: A parasocial relationship perspective. *Industrial Management & Data Systems*, *118*(3), 618–636. https://doi.org/10.1108/IMDS-05-2017-0214

Hasell, A. (2021). Shared Emotion: The Social Amplification of Partisan News on Twitter. *Digital Journalism*, *9*(8), 1085–1102. https://doi.org/10.1080/21670811.2020.1831937

Henkens, B., Verleye, K., & Larivière, B. (2020). The smarter, the better?! Customer well-being, engagement, and perceptions in smart service systems. *International Journal of Research in Marketing*, S0167811620300823. https://doi.org/10.1016/j.ijresmar.2020.09.006

Huffaker, D. (2010). Dimensions of Leadership and Social Influence in Online Communities. *Human Communication Research*, *36*(4), 593–617. https://doi.org/10.1111/j.1468-2958.2010.01390.x

Hui, Teo, & Lee. (2007). The Value of Privacy Assurance: An Exploratory Field Experiment. *MIS Quarterly*, *31*(1), 19. https://doi.org/10.2307/25148779

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), Article 1. https://doi.org/10.1609/icwsm.v8i1.14550

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language Style Matching Predicts Relationship Initiation and Stability. *Psychological Science*, *22*(1), 39–44. https://doi.org/10.1177/0956797610392928

Jang, Y.-T. J., Liu, A. Y., & Ke, W.-Y. (2022). Exploring smart retailing: Anthropomorphism in voice shopping of smart speaker. *Information Technology*

& People. https://doi.org/10.1108/ITP-07-2021-0536

Jiang, L., Zhou, W., Ren, Z., & Yang, Z. (2022). Make the apps stand out: Discoverability and perceived value are vital for adoption. *Journal of Research in Interactive Marketing*, *16*(4), 494–513. https://doi.org/10.1108/JRIM-03-2021-0076

Kehr, F., Kowatsch, T., Wentzel, D., & Fleisch, E. (2015). Blissfully ignorant: The effects of general privacy concerns, general institutional trust, and affect in the privacy calculus: Privacy calculus: dispositions and affect. *Information Systems Journal*, *25*(6), 607–635. https://doi.org/10.1111/isj.12062

Kowalczuk, P. (2018). Consumer acceptance of smart speakers: A mixed methods approach. *Journal of Research in Interactive Marketing*, *12*(4), 418–431. https://doi.org/10.1108/JRIM-01-2018-0022

Kudina, O., & Coeckelbergh, M. (2021). "Alexa, define empowerment": Voice assistants at home, appropriation and technoperformances. *Journal of Information, Communication and Ethics in Society*, *19*(2), 299–312. https://doi.org/10.1108/JICES-06-2020-0072

Lau, J., Zimmerman, B., & Schaub, F. (2018). Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 102:1-102:31. https://doi.org/10.1145/3274371

Lee, K., Lee, K. Y., & Sheehan, L. (2020). Hey Alexa! A Magic Spell of Social Glue?: Sharing a Smart Voice Assistant Speaker and Its Impact on Users' Perception of Group Harmony. *Information Systems Frontiers*, *22*(3), 563–583. https://doi.org/10.1007/s10796-019-09975-1

Li, Y., Mou, J., Ye, L., Long, J., & Huang, W. (Wayne). (2021). An empirical investigation of the utilitarian, social benefits in LBS information disclosure—The moderating effect of the gender based social role theory. *International Journal of Information Management*, *56*, 102243. https://doi.org/10.1016/j.ijinfomgt.2020.102243

Lucia-Palacios, Laura & Pérez-López, Raul. (2023). How can Autonomy improve consumer experience when interacting with smart products? 17(1), 19–37.

Ludwig, S., de Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing*, *77*(1), 87–103.

Maccario, G., & Naldi, M. (2022). Alexa, Is My Data Safe? The (Ir)relevance of Privacy in Smart Speakers Reviews. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2022.2058780

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, *15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Manikonda, L., Deotale, A., & Kambhampati, S. (2018). What's up with Privacy?: User Preferences and Privacy Concerns in Intelligent Personal Assistants. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 229–235. https://doi.org/10.1145/3278721.3278773

Massara, F. (2021). Unpacking the privacy paradox of consumers: A psychological perspective. *Psychology & Marketing*, 14. https://doi.org/10.1002/mar.21524

Novak, A. N., & Vilceanu, M. O. (2019). “The internet is not pleased”: Twitter and the 2017 Equifax data breach. *The Communication Review*, *22*(3), 196–221. https://doi.org/10.1080/10714421.2019.1651595

Park, C. S. (2015). Applying “Negativity Bias” to Twitter: Negative News on Twitter, Emotions, and Political Learning. *Journal of Information Technology & Politics*, *12*(4), 342–359. https://doi.org/10.1080/19331681.2015.1100225

Park, K., Kwak, C., Lee, J., & Ahn, J.-H. (2018). The effect of platform characteristics on the adoption of smart speakers: Empirical evidence in South Korea. *Telematics and Informatics*, *35*(8), 2118–2132. https://doi.org/10.1016/j.tele.2018.07.013

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, *71*(2001), 2001.

Petronio, S. (1991). Communication Boundary Management: A Theoretical Model of Managing Disclosure of Private Information Between Marital Couples. *Communication Theory*, *1*(4), 311–335. https://doi.org/10.1111/j.1468-2885.1991.tb00023.x

Phelps, J., Nowak, G., & Ferrell, E. (2000). Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*, *19*(1), 27–41. https://doi.org/10.1509/jppm.19.1.27.16941

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017). “Alexa is my new BFF”: Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’17*, 2853–2859. https://doi.org/10.1145/3027063.3053246

Ramadan, Z., F. Farah, M., & El Essrawi, L. (2020). From Amazon.com to Amazon.love: How Alexa is redefining companionship and interdependence for people with special needs. *Psychology and Marketing*, *May*. https://doi.org/10.1002/mar.21441

Scurlock, R., Dolsak, N., & Prakash, A. (2020). Recovering from Scandals: Twitter Coverage of Oxfam and Save the Children Scandals. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *31*(1), 94–110. https://doi.org/10.1007/s11266-019-00148-x

Septianto, F. (2020). Do past scandals influence the present performance? The moderating role of consumer mindset. *Journal of Business Research*, *106*, 75–81. https://doi.org/10.1016/j.jbusres.2019.09.011

Shapiro, D. L., Buttner, E. H., & Barry, B. (1994). Explanations: What Factors Enhance Their Perceived Adequacy? *Organizational Behavior and Human Decision Processes*, *58*(3), 346–368. https://doi.org/10.1006/obhd.1994.1041

She, J., Zhang, T., Chen, Q., Zhang, J., Fan, W., Wang, H., & Chang, Q. (2022). Which social media posts generate the most buzz? Evidence from WeChat. *Internet Research*, *32*(1), 273–291. https://doi.org/10.1108/INTR-12-2019-0534

Sheehan, K. B., & Hoy, M. G. (2000). Dimensions of Privacy Concern among Online Consumers. *Journal of Public Policy & Marketing*, *19*(1), 62–73. https://doi.org/10.1509/jppm.19.1.62.16949

Soroka, S., Daku, M., Hiaeshutter-Rice, D., Guggenheim, L., & Pasek, J. (2018). Negativity and Positivity Biases in Economic News Coverage: Traditional Versus Social Media. *Communication Research*, *45*(7), 1078–1098.

https://doi.org/10.1177/0093650217725870

Taylor, D. G., Davis, D. F., & Jillapalli, R. (2009). Privacy concern and online personalization: The moderating effects of information control and compensation. *Electronic Commerce Research*, *9*(3), 203–223. https://doi.org/10.1007/s10660-009-9036-2

Thakur, N., & Han, C. Y. (2018). An approach to analyze the social acceptance of virtual assistants by elderly people. *Proceedings of the 8th International Conference on the Internet of Things - IOT '18*, 1–6. https://doi.org/10.1145/3277593.3277616

Wang, C. L. (2021). New frontiers and future directions in interactive marketing: Inaugural Editorial. *Journal of Research in Interactive Marketing*, *15*(1), 1–9. https://doi.org/10.1108/JRIM-03-2021-270

Wang, X., & Lee, E. W. J. (2020). Negative emotions shape the diffusion of cancer tweets: Toward an integrated social network–text analytics approach. *Internet Research*, *31*(2), 401–418. https://doi.org/10.1108/INTR-04-2020-0181

Wang, Y., Han, R., Lehman, T. S., Lv, Q., & Mishra, S. (2022). Do Twitter users change their behavior after exposure to misinformation? An in-depth analysis. *Social Network Analysis and Mining*, *12*(1), 167. https://doi.org/10.1007/s13278-022-00992-8

Westin, A. F. (1967). Special report: Legal safeguards to insure privacy in a computer society. *Communications of the ACM*, *10*(9), 533–537. https://doi.org/10.1145/363566.363579

Wickham, H. (2016). tidyr: Easily tidy data with spread and gather functions.

Wickham, H., & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*.

Xu, J., & Wu, Y. (2015). Using Twitter in crisis management for organizations bearing different country-of-origin perceptions. *Journal of Communication Management*, *19*(3), 239–253. https://doi.org/10.1108/JCOM-06-2013-0050

Zhang, L., & Peng, T.-Q. (2015). Breadth, depth, and speed: Diffusion of advertising messages on microblogging sites. *Internet Research*, *25*(3), 453–470. https://doi.org/10.1108/IntR-01-2014-0021

Zhou, T. (2013). An empirical examination of user adoption of location-based services. *Electronic Commerce Research*, *13*(1), 25–39. https://doi.org/10.1007/s10660-013-9106-3

Zhu, X., Kim, Y., & Park, H. (2020). Do messages spread widely also diffuse fast? Examining the effects of message characteristics on information diffusion. *Computers in Human Behavior*, *103*, 37–47. https://doi.org/10.1016/j.chb.2019.09.006

Table 1. Effect of press releases on tweets´ sentiment (Model 1) and number of tweets (Model 2)

	Model 1 Sentiment		Model 2 Number of tweets	
Variables	**Beta**	**p-value**	**Beta**	**p-value**
VAPA				
Alexa	-		-	
Google	-0.61	<0.001	-538	<0.001
Siri	-0.33	0.006	-1,260	<0.001
N Positive news	-0.16	0.32	260	0.011
N Negative news	-0.30	0.003	213	<0.001
Lag dependent	0.07	0.21	-0.03	0.63
Year				
2019	-			
2020	0.52	<0.001	-579	<0.001
2021	0.34	0.017	-858	<0.001
Model fit	R^2= 0.188; Adj R^2= 0.169		R^2= 0.594; Adj R2= 0.584	

N= 306 (102 weeks and 3 VAPAS)

Table 2. Differences in sentiment between privacy and non-privacy tweets.

Variable	General, N = 432,186[1]	Privacy, N = 9,341[1]	p-value[2]
Affect (Liwc)	5.01 (5.87)	3.70 (4.63)	<0.001
Tone (Liwc)	49.54 (37.27)	40.07 (34.35)	<0.001
Risk (Liwc)	0.16 (1.04)	0.69 (2.22)	<0.001
Compound (Vader)	0.22 (0.41)	0.14 (0.41)	<0.001

[1] Mean (SD), [2] Wilcoxon rank sum test

Table 3. Effect of mentions of different keywords on the sentiment (Model 3.1) of the tweets and number of tweets (Model 3.2).

Variables	**Model 3.1. Sentiment**		**Model 3.2 Number of tweets**	
	Beta	p-value	Beta	p-value
VAPA				
Alexa	—			
Google	0.23	0.11	0.09	0.7
Siri	0.08	0.5	0.21	0.4
Mentions record	-1.2	<0.001	1.2	<0.001
Mentions voice	-0.52	<0.001	0.40	<0.001
Mentions data	-0.64	<0.001	1.1	<0.001
Mentions listen	-1.1	<0.001	0.20	0.012
Mentions user	0.01	>0.9	0.80	<0.001
Mentions conversation	-1.0	<0.001	0.43	0.012
Mentions concern	-1.7	<0.001	0.71	<0.001
Mentions contractor	-0.68	0.005	-0.28	<0.001
Mentions home	0.39	0.032	0.05	0.6
Mentions device	0.09	0.6	1.6	<0.001
Lag dependent			-0.01	0.3
Year				
2019	—		—	
2020	-0.36	0.005	0.10	0.7
2021	-0.56	0.002	0.13	0.6
Model fit	R2 = 0.046; Adj R2 = 0.045; nobs = 7,994		R2 = 0.951; Adj R2 = 0.951; nobs = 1,509	

Table 4. General discussion of main findings and managerial implications.

Main Findings	Discussion	Managerial Implications
Privacy tweets are more negative and include more “risk” words than those not speaking about privacy, which score higher in emotional tone and include a lower number of “risk” words. The sentiment of VAPAs’ privacy-related tweets becomes more negative over time.	These results corroborate previous literature, which highlights that consumers are usually concerned with the risks associated to online privacy and VAPAS, and that they are skeptical with how companies use private data (Hsieh & Lee, 2021; Kowalczuk, 2018; Malhotra et al., 2004; Taylor et al., 2009) These findings are also in line with Cao and Wang (2022) and–Massara (2021),_who evidence that, with the advent of AI and technological advances, consumers are more worried than before about the use of private data. Besides, as highlighted by Kudina and Coeckelbergh (2021), since many scandals related to the bad use of private data by companies have come to light, consumers do not trust too much in companies. Unlike most previous literature that relies on survey data, this research uses Twitter data. One of the main advantages of using social buzz instead of surveys is that consumers, in general, post opinions on social media without direct prompting or influence from marketers. Therefore, consumers are more likely to express their true perceptions about products, services or brands and to inform about their true behaviors (Kozinets et al., 2010). Although some studies include perceived VAPAs privacy and trust on VAPAs as an antecedent of adoption or continuance use of VAPAs in their conceptual models, using theories such as TAM and UTAUT2 (Hsieh & Lee, 2021; Kowalczuk, 2018), there has been a limited research on the evolution of consumers´ sentiment related to privacy of VAPAs over time. Several factors such as companies scandals or new functionalities added to VAPAs might change the sentiment of that conversation, which might not be always the same.	Companies should proactively respond to evolving consumer sentiment and privacy concerns related to VAPAs by trying to implement privacy policies that meet consumers expectations. Continued monitoring of VAPAs' privacy concerns and regular updates to privacy policies may be necessary to avoid dissatisfied consumers generating bad comments about them.
Negative and positive news increase VAPAs buzz, but negative content has a stronger effect. Negative news lead to a more negative sentiment of tweets, while positive news	We contribute to the social buzz literature by confirming that Twitter is a very popular place to discuss news about companies (Novak & Vilceanu, 2019), since press releases about VAPAs increased social buzz in Twitter. This result is in line with that from ~~(~~She et al. (2022), who found that the more emotional the language is (no matter whether positive or negative), the more attractive the message is. This attractiveness might be evidenced in our research by the increase in consumers social buzz. Besides, we can also corroborate that negative	Firms must strengthen security measures to avoid reputational crises. Otherwise, consumers might feel betrayed when scandals about bad use of personal data come into light.

does not impact the sentiment of social buzz.	content of news might have a stronger effect on consumers than positive content (Park, 2015; Zhu et al., 2020). In fact, negative content is even more attractive than positive content (She et al., 2022).	
Apple Siri is the VAPA with the highest percentage of privacy-related tweets.	While many studies highlight the benefits of transparent communication about data use (Ashworth & Free, 2006; Hsieh & Lee, 2021; Malhotra et al., 2004), our findings indicate that even brands that emphasize their privacy focus, like Apple's Siri, can face intensified scrutiny and skepticism. This disparity emphasizes the complexities inherent in managing and communicating about privacy in today's digital age.	Even brands emphasizing privacy (like Apple) can face intensified scrutiny and skepticism. A possible explanation might be that consumers have higher expectations in terms of privacy protection and, when they are not met, dissatisfaction occurs. Therefore, companies should be careful on which they are promising. Generating high expectations might lead to dissatisfied consumers when these expectations are not met. It might be better to only promise what they actually can offer.
Terms aligned to the feeling of being spied and listened, such as like "concern," "record," "listen," and "conversation", generate more social buzz and predict negative sentiment in privacy-related tweets.	These findings are in line with Mani and Chouk (2019), who reveal that the feel of being spied and listened, the lack of control of personal data and, the unauthorized use of personal information, are drivers of resistance to use smart services. These relevant terms raised by consumers when speaking about privacy concerns are also aligned with Malhotra et al. (2004), who highlighted that consumers´ online privacy concerns depend on the extent to which an internet user is concerned about the gathering of their personal information by online marketers, their capacity to control that information, and their awareness of how it is being utilized. It also corroborates the findings of Lau et al. (2018) and Tabassum et al. (2019), who claimed that consumers'' primary concern revolved around the potential of smart speakers to listen and record audio, which evidence a lack of control over them, which is also the basis of the social exchange theory (Ashworth & Free, 2006; Chen et al., 2022; Malhotra et al., 2004).	Companies should address issues related to surveillance, data collection, unauthorized access, and loss of personal information. From the privacy calculus theory perspective, it is important to ensure that the benefits and the risks of information sharing are transparently communicated. Otherwise, consumers feel betrayed when things do not happen as they expect. From the perspective of social contract theory, awareness of privacy policies is important, so public managers can play an independent role in providing tools that allow consumers to better understand these privacy policies, as well as communication policies about them in the press. In any case, the implementation of these public policies will largely depend on local legal and cultural contexts.

Appendix A

Most frequent words in privacy-related tweets