Reciis – Electronic Journal of Communication, Information & Innovation in Health, Rio de Janeiro, v. 18, n. 4, p. 98-113, Jan.-Mar, 2024
[www.reciis.icict.fiocruz.br] e-ISSN 1981-6278

https://doi.org/10.29397/reciis.v18i1.3859

ORIGINAL ARTICLES

Videos about vaccines: what factors influence on having higher views on YouTube?

Vídeos sobre vacinas: quais fatores influenciam em maior visualização no YouTube?

Videos sobre vacunas: ¿qué factores influyen en una mayor visualización en YouTube?

Arthur da Silva Lopes[1,a]
arthur.lopes@ufba.br | https://orcid.org/0000-0001-9137-3184

Antonio Marcos Pereira Brotas[2,b]
brotas@bahia.fiocruz.br | http://orcid.org/0000-0001-8438-2445

[1] Federal University of Bahia, Collective Health Institute. Salvador, BA, Brazil.
[2] Oswaldo Cruz Foundation, Gonçalo Moniz Research Center. Salvador, BA, Brazil.

[a] Undergraduate Degree in Health from the Federal University of Bahia.
[b] Ph.D. Program in Culture and Society by the Federal University of Bahia.

ABSTRACT

Considering the growing importance of YouTube as a source for health information search, the aim of this study was to analyze the factors associated with a higher number of views in videos about covid-19 vaccines. For this purpose, Natural Language Processing techniques and statistical modeling were employed based on 13,619 videos, encompassing three types of variables: general metrics, textual content of titles, and information about the participants in the videos. Among the results, videos of medium or long duration, posted during late hours and on weekends, with tags, descriptions, and short titles, along with controversial elements and the presence of male and white figures in thumbnails stand out. These findings contribute to a better understanding of the potential factors to be considered in the production of health communication content about vaccines on YouTube.

Keywords: Health communication; Health education; YouTube; Vaccines; Engagement.

RESUMO

Considerando-se a crescente importância do YouTube como fonte para busca de informações em saúde, o objetivo deste trabalho é analisar os fatores associados a um maior número de visualizações de vídeos sobre vacinas contra a covid-19. Para isso, usaram-se técnicas de Processamento de Linguagem Natural e modelagem estatística com base em 13.619 vídeos, abrangendo três tipos de variáveis: métricas gerais, conteúdo textual dos títulos e informações sobre os participantes dos vídeos. Entre os resultados, destacam-se os vídeos de duração média ou longa, postados durante a madrugada e nos fins de semana, com *tags*, descrição e títulos curtos, além de elementos controversos e presença de figuras masculinas e brancas em miniaturas. Os achados contribuem para uma melhor compreensão dos possíveis fatores a serem considerados na produção de conteúdo de comunicação em saúde sobre vacinas no YouTube.

Palavras-chave: Comunicação em saúde; Educação em saúde; YouTube; Vacinas; Engajamento.

RESUMEN

Teniendo en cuenta la creciente importancia de YouTube como fuente de búsqueda de información en salud, el objetivo de este artículo es analizar los factores asociados a un mayor número de visualizaciones en videos sobre vacunas contra el covid-19. Para eso, se emplearon técnicas de Procesamiento del Lenguaje Natural y modelado estadístico basadas en 13,619 videos, que abarcan tres tipos de variables: métricas generales, contenido textual de títulos y información sobre los participantes en los videos. Entre los resultados, destacan los videos de duración media o larga, publicados durante altas horas de la noche y los fines de semana, con *tags*, descripciones y títulos cortos, junto con elementos controvertidos y la presencia de figuras masculinas y blancas en las miniaturas. Estos hallazgos contribuyen a una mejor comprensión de los posibles factores a tener en cuenta en la producción de contenido de comunicación de salud sobre vacunas en YouTube.

Palabras clave: Comunicación en salud; Educación en salud; YouTube; Vacunas; Interacción.

ARTICLE INFORMATION

Authors' contributions:
Conception and design of the study: Arthur da Silva Lopes.
Acquisition, analysis or interpretation of data: Arthur da Silva Lopes.
Manuscript writing: Arthur da Silva Lopes.
Critical review of intellectual content: Antonio Marcos Pereira Brotas.

Declaration of conflict of interests: none.

Sources of financing: none.

Ethical considerations: none.

Additional acknowledgments/Contributions: none.

Article history: submitted: 28 Jun. 2023 | accepted: 29 Aug. 2023 | published: 28 Mar. 2024.

Previous presentation: none.

INTRODUCTION

André Pereira Neto and collaborators (2015) point out that the internet can be considered the most emblematic Information and Communication Technology (ICT) of contemporary times. It has caused substantial changes in the ways individuals inform, communicate, and relate to each other. The internet has reconfigured the roles of those who produce and those who consume information, broken geographic boundaries, and made ubiquitous and instantaneous access and sharing.

Regarding the impact of these transformations in the field of health, those inserted in the scope of communication and education must be highlighted, especially due to what McClung, Murray, and Heitlinger pointed out at the end of the 20th century: "[...] the public wants access to information about medical problems. Information providers on the internet (net) responded to this desire, and medical information is abundant" (1998, p. 1, our translation).

At that time, this information was limited to search engines and websites. Currently, people can find health establishments and professionals, access other people's evaluations of care on medical websites, find out about therapeutic and diagnostic possibilities, and discuss among peers in forums and socio-digital networks (Garbin; Pereira Neto; Guilam, 2008).

HEALTH INFORMATION ON SOCIAL MEDIA PLATFORMS: THE CASE OF YOUTUBE

The Digital News Report (Newman, 2022) – a report published by the Reuters agency – points out that information sources are found notably in online environments (83%), with YouTube being the most used for news consumption (43%), and the second for general uses (77%), only behind WhatsApp (78%). It is for this reason that YouTube is treated here as a locus of investigation, and the videos produced on this platform are our object of analysis.

Despite the availability and abundance of information, what people consume is not always reliable. When analyzing 242 videos in Portuguese about acute myocardial infarction, Fialho *et al.* (2021) identified that a third of those videos were irrelevant and contained inaccurate data, and among the relevant ones, the average DISCERN[1] quality was low or moderate.

In a more recent study carried out by Dalpoz *et al.* (2022), 48 educational videos on preventing dental caries were investigated; the analysis showed that the majority of videos were of moderate quality according to the DISCERN criteria and their reliability (JAMA) was in general low due to a lack of transparency regarding the origin of the information presented.

Chan and collaborators (2021) analyzed videos about vaccines against covid-19. Most videos were considered non-educational, with "fair" average DISCERN quality. The educational videos had quality ranging from "fair" to "excellent," depending on the origin of the producer, and the information reliability (HONcode[2]) was "low" for most non-educational videos and "moderate" for the educational ones. Overall, these authors emphasize that the quality and reliability of the information are far from ideal.

The dialectic between quality and quantity has certainly intensified in the context of infodemic and information disorder (Wardle; Derakhshan, 2017). Concerning vaccines, the scenario appears to be conducive to the action of anti-vaccine movements that are important agents in promoting vaccine hesitancy (Pierri *et al.*, 2022), which we know is a relevant public health problem today (WHO, 2019).

Faced with this scenario, researchers have focused on the growth of misinformation about vaccines on YouTube (Brotas; Costa; Massarani, 2021; Calvo; Cano-Orón; Llorca-Abad, 2022; Massarani *et al.*, 2021).

---

[1] Instrument used to evaluate the quality of health-related information.

[2] This is a certification used to evaluate the quality of health-related information, by verifying the origin, transparency, and intentionality (ethical dimension) of the content produced or made available on the internet.

Aiming at identifying the occurrence of disinformative content and understanding in which situations it occurs, these studies were qualitative for the most part and took the selected most relevant videos as objects of analysis.

It is important to emphasize that this "relevance" is given according to the criteria of YouTube itself, which, belonging to a private company, Google, seeks the greatest number of interactions (Souza, 2020). It is a recommendation system that works based on the footprints left by users when using the platform. Although these are the videos most likely to be accessed and become sources of information, they do not represent all the productions available on YouTube on the topic.

Despite the importance and relevance of qualitative studies, there is still a lack of quantitative studies that aim to investigate which factors are associated with a greater number of views of videos about vaccines on YouTube, considering a comprehensive set of possible explanatory variables.

It is important to highlight that there may be other productions of better quality, but that 'for some reason' they were unable to gain enough audience to compete for relevance. The literature on the subject still does not understand what leads certain videos about vaccines to have more audiences than others. Thus, it is essential to understand the principles that guide the production of this content – whether they are reliable or not – and highlight possible criteria used by recommendation systems that favor information disorder.

As Li and collaborators (2020) highlight, YouTube has the potential to be an opportune tool for health communication in situations of health crises, such as the covid-19 pandemic, if it is properly used by public institutions and healthcare professionals to disseminate reliable and quality information. Therefore, the authors recommend that "[...] these groups should find strategies to increase the number of views and the impact of their videos [...]" on the platform (Li *et al.*, 2020, p. 5, our translation).

The general objective of this study was to identify the factors associated with a greater number of views for videos about vaccines on YouTube in order to fill the gap in the literature on the issue. This information can guide health and science communication professionals and institutions so that they can compete for larger audiences, and thus provide reliable information and combat misinformation.

METHODOLOGY

The quantitative research approach was used to respond to the research objective. The unit of analysis will not be the videos with the highest engagement (as in the qualitative studies discussed previously), but all those posted on the platform between 2020 and 2021, containing the word "vaccine" in the title or description, voluntarily published by users and subject to be collected through the YouTube Data Tools platform (Rieder, 2023).

A total of 13,619 videos resulted from the collection procedure. Despite not dealing with a health issue, Munaro *et al.* (2021) and Yang *et al.* (2022) carried out statistical studies on the factors that influence toward greater number of views and other metrics on YouTube. The independent variables are comprehensive, and those identified by these studies as having a statistically significant effect, when applicable, on the number of views (dependent variable of this study) were selected for this study and are detailed in Table 1.

Table 1 – Description of categorical variables related to general metrics

Categorical variable	Annotation	Category	N.	Proportion (%)
Posting (part of the day)	part_of_day	evening	5,280	38.76
		afternoon	4,595	33.73
		night	2,009	14.75
		morning	1,735	12.73
Posting (day of the week)	day_of_week	Monday (Mon)	2,640	19.38
		Tuesday (Tue)	2,586	18.98
		Wednesday (Wed)	2,542	18.66
		Thursday (Thur)	2,396	17.59
		Friday (Fri)	2,129	15.63
		Saturday (Sat)	836	6.13
		Sunday (Sun)	490	3.59
Duration of video	duration	short (< 10 min)	11,710	85.98
		medium (10-19:59 min)	1,085	79.66
		long (> 20 min)	824	6.05
Number of words in the title	len_video_title	over (>=12 words)	6,588	48.37
		under (<= 10 words)	5,380	39.50
		medium (10-12 words)	1,651	12.12
Number of uppercase characters	n_upper	under (<= 6 characters)	10,428	76.65
		over (>= 8 characters)	2,739	20.11
		medium (6-8)	452	3.31
Use of description	desc_use	1 (Yes)	12,874	94.52
		0 (No)	745	5.47

Source: Prepared by the authors.

VARIABLES ASSOCIATED WITH THE TEXTUAL CONTENT OF THE TITLES

For Named Entity Recognition (NER) in titles, we used our algorithm, the ProfileNER-classifier[3]. NER is a technique belonging to the area of Natural Language Processing (NLP) that comprises the identification and classification of entities, which can be any set of words that refer to a certain category.

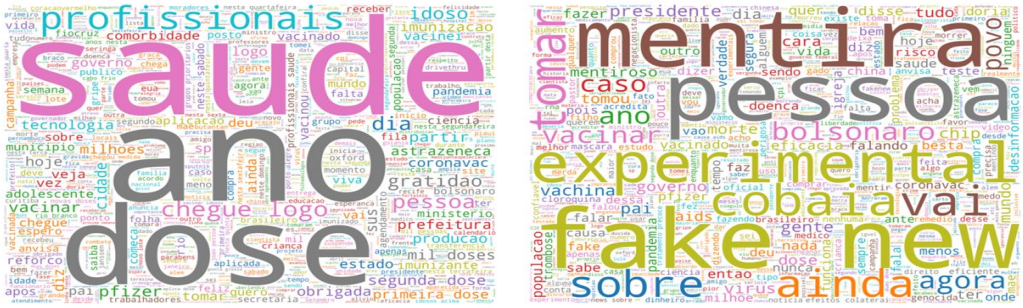
Thus, headache, cough, muscle pain, and fever are terms that make up the entity 'symptoms,' for example. Among the entities used in this study, we have: "vaccines," "political vaccines" (vac_politics), "products," "medicines," "diseases," "science," "body parts," "frequent events" (eventos_freq), and "rare events" (eventos_raros). The variable related to NER corresponds to the number of occurrences of each entity.

The entity "vaccines" refers to the names of the laboratories that produce them (e.g. AstraZeneca, Oxford, Butantan, Fiocruz) while "political vaccines" refers to the names: vachina, Doria vaccine, China vaccine, Xing vaccine, etc.. The entity "products" refers, among others, to respirators, syringes, masks, PPE, PCR, and hospital beds.

For the task of automated detection of potentially uninformative or controversial content in titles, an initially trained machine learning model was used, as explained by Lopes and Brotas (2023). Throughout this investigation, there was an investment in improving the model's performance, particularly in terms of

[3] Algorithm whose operation takes place through dictionaries of terms related to each entity when in the NER function, or related to each profile. These dictionaries are available in the Profileclassifier folder.

attribute selection and engineering, and an increase in training data (textual content illustrated in Figure 1) through the collection and classification of tweets.



(a) dados classificados como "outros" (b) dados classificados como "controversos"

Figure 1 – Cloud of words present in the model training data
Source: Prepared by the authors.

The choice for Twitter was due to two aspects: 1) suitability for the task to be performed by the model; and 2) efficiency. Regarding the first, Twitter was effectively the stage for debates and disputes around vaccines against covid-19 (Penteado *et al.*, 2021). Regarding the second, the platform is a space through which people and institutions share information, opinions, news, daily reports, etc in relatively short texts (280 characters), reducing operational costs (from manual reading for validation) and computational costs (in model training).

The descriptors used in the collection corresponded to the assigned thematic categories, and validation consisted of a manual reading of all tweets, except for those belonging to the themes "security" and "named disinformation" indicated in a sample of 50% of the tweets with these themes. The result is summarized in Chart 1.

Chart 1 – Themes of the new training data incorporated into the model

Theme (n.)	Tweet-example
Safety (1,200)	Anvisa MUST, before recommending, present robust studies proving the efficacy and safety of the vaccine. Offering a 'vaccine' without proof of safety is GENOCIDE.
Expired vaccines (300)	And does it make any difference to use an expired "vaccine"? Does a biological weapon have an expiration date?
Vaccine as an object of profit (300)	After Pfizer made trillions in profits from a very questionable vaccine, and after the social control experiment was concluded, we declared the end of the pandemic.
Adverse effects of the vaccine (300)	Has this woman gone completely crazy? WHO declaring the end of the pandemic, countries suspending vaccination due to adverse effects, studies coming to light and concluding that the vaccine causes myocarditis, and she is talking about "intensifying"!? 🤨
Vaccine Adverse Event Reporting System (VAERS) (600)	She died from this vaccine, yes, TTP is one of the reactions of these vaccines, they are lying through their teeth, a 5-month-old baby died in the United States after his mother took the vaccine, this is in VAERS.
Named disinformation (false, misleading, and lying content) (1,843)	False publications about covid-19 vaccines grew 383% in 2 months. Anti-vaccine groups use social media to echo false content and generate distrust about the future vaccination campaign.

Source: Prepared by the authors.

Hence, there was an improvement in all performance metrics: accuracy of test data not yet seen by the model (92% to 94%); a macro average of the F1-score (harmonic mean between precision and sensitivity, 92% to 94%); precision (92% to 95%); and sensitivity (92% to 93%), showing better generalization capacity.

VARIABLES ASSOCIATED WITH THE VIDEOS' MEDIATOR (THE ONE WHO APPEARS IN THE VIDEOS)

The presence of human faces was detected in the videos' *thumbnails* under the hypothesis that the presence of the mediators themselves on the 'cover' of the videos is a possible strategy for mobilizing their social capital (Recuero; Soares, 2021).

The gender and race classification was carried out based on the videos' *thumbnails* so that it was possible to investigate whether orders resulting from these analytical-structural categories also affected the number of views by using the machine and deep learning model and VGG-Face (accuracy of 89.3%, F1 = 88.3%) through the DeepFace framework (Serengil; Ozpinar, 2020).

As discussed by Butler (2003), sex itself would be a sociocultural construct because it is possible to go through its genealogy and find assigned aspects and meanings that are different from those considered hegemonic (biomedical). However, the variable named gender as sex, in its dysmorphic-anatomical conception (on which the model used was trained) is understood as a constituent component of (cis)gender intelligibility.

The model shows 97% accuracy. Most were assigned to the "none" category, i.e., those in which the algorithm was unable to detect the presence of a human face; therefore, these occurrences were disregarded, as were those in which it was not possible to distinguish between "man" or "woman".

Regarding race/color, the theoretical assumption that supported adoption through automated classification was that of Oracy Nogueira (2007) who points out that race prejudice in Brazil is mediated (and produced) by the mark, meaning by the phenotype. However, the complexity of Brazilian racism, and the very notion of race, evidenced by the disputes surrounding public affirmative action policies (Nunes, 2018), presents itself as a challenge to the attempt at automated classification, as initially proposed by the authors of DeepFace and the database used for training (Kärkkäinen; Joo, 2019).

For this reason, the model's categories – based on the US census: white, black, Latin American, Indian, Arab, and Asian – were adapted, and the initially multiclass classification became binary with the variables of white and non-white. This was due to the theoretical unfeasibility of simply importing a racial interpretation resulting from a divergent social manifestation mechanism (of origin in the case of the United States and according to Nogueira) and the objective of this investigation being restricted to the identification of possible asymmetries associated with different socialization processes mediated by race (racialization) of phenotypically white individuals and those who diverge from this norm.

The classification procedure for race/color was also changed. As the model shows 68% accuracy, different image versions[4] were used to capture multiple thumbnail frames (as most of them are dynamic, that is, they show fragments of the videos).

Thus, a script capable of detecting race/color was created and applied to each of the frames (thus overcoming possible biases associated with lighting or the person's position), and thus the results were stored in a variable of the list type. Once this step was completed, the final classification was based on the mode (using Python's Statistics module, which indicates the most frequent value in a given set of data) of this list. Descriptive statistics for the presence of a human face, gender, and race/color are shown in Table 2.

[4] Access to video thumbnails on YouTube is based on the following structure: https://img.youtube.com/vi/<inserir-id-do-video-aqui>/inserir-formatos.jpg; the formats used here are: 0.jpg, hq1.jpg, hq2.jpg and hq3.jpg.

Table 2 – Description of the categorical variables associated with the video mediator.

Categorical variable	Annotation	Category	N.	Proportion (%)
Face in thumbnail	Face	0 (No)	7,958	58.43
		1 (Yes)	5,661	41.56
Race/ethnicity*	Race/ ethnicity	0 (Non-white)	163	3.01
		1 (White)	5,250	96.99
Gender*	Gender	0 (Woman)	1,503	27.77
		1 (Man)	3,910	72.23

Source: Prepared by the authors.
* For "yes" face. The total is equivalent to 5,413, as the videos in which the race script could not reach a result were subtracted from 5,661.

The analysis of the correlation between the variables was performed using negative binomial regression as reported by Munaro *et al.* (2021) and Yang *et al.* (2022) due to the nature of the dependent variable being countable and presenting overdispersion.

RESULTS

Therefore, the results regarding the platform's native variables are shown in Table 3. Most of the videos had a description (desc_use) that showed a positive association with a greater number of views on the platform. Regarding the length of the video, most of those about vaccines were short (< 10 min.), however, those of medium (β = 1.22) and long (β = 1.02) duration were most associated with the highest number of views.

Table 3 – Description of the model's results regarding the variables associated with the platform.

(continue)

Category	Coefficient	p-value ($\alpha < 0.05$)	Confidence interval (95%)	
intercept	2.5717	0.000	2.520	2.623
desc_use	2.1008	0.000	2.025	2.177
duration_short	0.3336	0.000	0.299	0.368
duration_long	1.0191	0.000	0.964	1.074
duration_medium	1.2190	0.000	1.169	1.269
n_upper2_under	0.9697	0.000	0.932	1.007
n_upper2_over	1.0209	0.000	0.978	1.063
n_upper2_medium	0.5811	0.000	0.512	0.650
len_video_title2_ under	0.9239	0.000	0.894	0.954
len_video_title2_over	0.7911	0.000	0.761	0.821
len_video_title2_medium	0.8567	0.000	0.817	0.896
len_tags2_ under	0.9026	0.000	0.872	0.934
len_tags2_over	1.6691	0.000	1.636	1.702
part_of_day_night	1.1820	0.000	1.141	1.223
part_of_day_morning	0.3798	0.000	0.339	0.421
part_of_day_evening	0.5117	0.000	0.482	0.541
part_of_day_afternoon	0.4982	0.000	0.468	0.528
videoCategoryLabel_Comedy	1.3932	0.000	0.996	1.791
videoCategoryLabel_Education	1.3675	0.000	1.221	1.514

(conclusion)

videoCategoryLabel_ Entertainment	-0.0446	0.506	-0.176	0.087
videoCategoryLabel_Film	-2.0613	0.000	-2.409	-1.713
videoCategoryLabel_Gaming	-2.2304	0.000	-3.282	-1.178
videoCategoryLabel_How to	1.3027	0.000	0.950	1.656
videoCategoryLabel_Music	4.6596	0.000	4.347	4.972
videoCategoryLabel_News	0.2837	0.000	0.157	0.410
videoCategoryLabel_Nonprofit	0.5720	0.000	0.293	0.851
videoCategoryLabel_People	-0.2379	0.001	-0.373	-0.103
videoCategoryLabel_Science	1.5568	0.000	1.407	1.707
videoCategoryLabel_Sports	0.1173	0.453	-0.189	0.424
videoCategoryLabel_Travel	-4.1069	0.000	-4.928	-3.286
day_of_week_Sun	0.4412	0.000	0.362	0.520
day_of_week_Wed	0.1393	0.000	0.100	0.179
day_of_week_Thur	0.1582	0.000	0.120	0.197
day_of_week_Sat	0.5492	0.000	0.487	0.612
day_of_week_Mon	0.4288	0.000	0.386	0.471
day_of_week_Fri	0.5040	0.000	0.464	0.544
day_of_week_Tue	0.3510	0.000	0.312	0.390

Source: Prepared by the authors.

Regarding the number of words in the titles, there was a greater occurrence of videos above the average (>= 12 words) but those with 10 words or less (β = 0.92) were viewed more. Most videos used 6 or fewer capitalized characters in their titles, however, those that contained at least 8 capitalized characters (β = 1.02) were associated with a greater number of views even though they represented only 20.11% of the sample.

Regarding the weekday when videos were posted, Saturday and Sunday were the days on which channel mediators posted the least. However, the days most positively associated with greater viewing were Friday (β = 0.50), Saturday (β = 0.55), Sunday (β = 0.44), and Monday (β = 0.43); that is, at the end and beginning of the week. When analyzing what part of the day the video was posted, the majority were posted in the evening or the afternoon, although at night (during the night) (β = 1.18) was the time associated with the most views.

Finally, videos assigned to the "music" category (β = 4.66) were associated with a greater number of views, followed by "science" (β = 1.56), "comedy" (β = 1.39), "education" (β = 1.37), and "how-to" (β = 1.30). Regarding the use of the YouTube tag tool, videos in which mediators used at least 11 tags (β = 1.66) showed a higher number of views.

Out of the variables linked to the textual content of the video titles, as described in Table 4, a positive association was observed between the presence of elements linked to misinformation/controversy about vaccines (β = 1.23) and a greater number of views as well as the use of terms associated with etiological agents (β = 1.08), post-vaccination events (β = 0.13 for frequent ones, β = 0.63 for serious ones), diseases (β = 0. 89), and products (β = 0.24).

Table 4 – Model results for variables at the textual level of video titles

Category	Coefficient	Standard deviation	p-value ($\alpha < 0.05$)	Confidence interval (95%)	
intercept	8.7165	0.016	0.000	8.685	8.748
diseases	0.8690	0.021	0.000	0.829	0.909
parts of the body	-0.1645	0.014	0.000	-0.193	-0.136
vaccine	-0.0697	0.020	0.001	-0.110	-0.030
products	0.2367	0.070	0.001	0.100	0.373
science	-0.1878	0.050	0.000	-0.286	-0.090
medicines	0.5454	0.119	0.000	0.311	0.779
events_freq	0.1765	0.038	0.000	0.102	0.251
events_serious	0.6253	0.122	0.000	0.386	0.864
agents	1.0776	0.039	0.000	1.001	1.154
vac_politics	-0.0701	0.071	0.320	-0.208	0.068
controversy	1.2327	0.019	0.000	1.195	0.032

Source: Prepared by the authors.

When analyzing the results of the variables' regression referring to the characteristics related to the mediator of the videos, as shown in Table 5, it is observed that the use of human figures in the *thumbnail* of the videos ($\beta = 0.37$) was positively associated with a greater number of views. These figures – of men ($\beta = 0.85$) and whites ($\beta = 0.92$) – were those that most positively corroborated a greater number of views on the platform, while the counterparts showed a negative correlation.

Table 5 – Description of the model results for the variables associated with the mediator

Variable	Coefficient	p-value ($\alpha < 0.05$)	Confidence interval (95%)	
intercept [homem/white]**	8.2760	0.000	8.114	8.438
intercept [mulher/non white]**	10.0464	0.000	10.015	10.078
gender [man]	0.8475	0.000	0.788	0.907
gender [woman]	-0.8475	0.000	-1.079	-0.767
race/skin color [non-white]	-0.9229	0.000	-0.907	-0.788
race/skin color [white]	0.9229	0.000	0.767	1.079
human figure***	0.3675	0.000	0.333	0.402

Source: Prepared by the authors.
**Different bivariate regressions. Although the race/color and gender variables are binary and internally correlated, leading to multicollinearity, the coefficients were presented to highlight the positive and negative effects on the predictor variable.
*** Univariate regression.

DISCUSSION

Although not directly focused on vaccines, studies such as Yang *et al.* (2022) clarify some of the factors related to greater engagement on YouTube. Analyzing the Reactions channel (chemistry scientific dissemination channel), the authors indicate a positive association between shorter and more concise titles and a greater number of views, corresponding to the findings of the present study. The hypothesis pointed out by

the authors was that this result may be a consequence of the fact that, when longer titles are "cut" on the platform, that is, when they are not presented in full to users, they do not translate their content.

Conversely, Yang and collaborators (2022) point out that the length of the videos had a negative coefficient in relation to the number of views, and, for this reason, the shorter videos were the most watched. This result was also found by Munaro *et al.* (2021), however, it does not match the findings of this research showing that medium and long videos presented considerably higher coefficients than those referring to short videos.

The explanation for this divergence may be related to the themes covered in the analyzed videos and the socio-historical moments of production and publication. The hypothesis suggested here is that for videos about vaccines published in the context of covid-19, the results indicate that people who searched for information about immunizations on YouTube tend to favor videos with greater explanatory potential unlike the results found by Munaro and collaborators in the aforementioned study.

Faced with a health event of the magnitude of the covid-19 pandemic, marked by all sorts of uncertainties (Lima *et al.*, 2020) – by widespread politicization, especially regarding vaccines in Brazil, and by informational disorder (Recuero *et al.*, 2021) –, it is suggested that the perception of risk has challenged information consumption behavior on the platform, associated with the fact that research shows that YouTube's recommendation system favors long videos over shorter ones (Calvo; Cano -Orón; Llorca-Abad, 2022).

These circumstances are highlighted here as hypotheses so that, to the same extent that videos were sought that supposedly could "pass" more information, there was also a greater inclination to view videos that contained some element associated with misinformation or controversy in their titles. In addition, they mention terms related to post-vaccination events, whether serious (thrombosis, anaphylaxis, myocarditis, pericarditis, thrombocytopenia, etc.) or frequent (itching, fever, cough, tiredness, fatigue); or terms related to diseases (especially "covid"), etiological agents (especially "viruses"), and medications (hydroxychloroquine, drugs, early treatment, azithromycin, etc.). There is also a greater number of characters in capital letters – already associated with misinformation (Damstra *et al.*, 2021).

Although these results imply a reality that is conducive to informational disorder, the fact that videos included in the "science" category are associated with a greater number of views indicates that these elements were also used by those who aimed to practice disinformation, such as scientific communicators and non-specialized YouTubers.

In fact, this is confirmed when analyzing the four videos with the highest number of views among the 13,619 that make up the analyzed sample of this study: "Interview with Bolsonaro: betrayal, elections, vaccine, economy, and much more"; "COVID VACCINE: WHAT THEY DIDN'T TELL YOU ABOUT THE THIRD DOSE"; "Is the covid-19 vaccine dangerous? Anthony Wong responds;" and "COVID: OXFORD ASTRAZENECA VACCINE CAUSES THROMBOSIS???". Those in capital letters belong to the channel Olá, Ciência!, an important vehicle for scientific dissemination about vaccines on YouTube whose relevance has already been highlighted in other studies (Massarani; Costa; Brotas, 2020).

However, authors such as van der Linden (2022) point out that the strategy of confronting misinformation or dealing with controversies by reproducing all or parts of the content that is intended to counteract (e.g. emphasizing that the Oxford vaccine does not cause thrombosis in a suggestive tone) can lead to an opposite effect, that is, instead of mitigating and suppressing, it ends up maintaining and propagating these terms/narratives across the network, contributing with the problems of sensationalism and clickbait. In any case, this strategy seems to give science-focused channels better conditions to compete for audience with those who spread misinformation.

Regarding this competition, when analyzing the interactions between pro-vaccine and anti-vaccine groups on Facebook, Johnson and collaborators (2020) show that, despite being a minority, anti-vaccine groups demonstrated to be more persuasive as they managed to co-opt a greater number of undecided users into their cluster. One of the reasons given is that, when the content shared in the different groups was analyzed, it was noticed that those belonging to anti-vaccine groups were more plural, while that produced and shared by pro-vaccine groups ended up being more monothematic.

Furthermore, Munaro *et al.* (2021) point out a positive association between videos posted during non-business hours, and on Mondays, Tuesdays, or Thursdays, and a greater number of views. In fact, videos published in the early hours of the morning and on Mondays positively influenced the number of views. However, according to our results, Thursday was one of the worst days to upload videos about vaccines in the context of the covid-19 pandemic; with Monday, Friday, Saturday, and Sunday being those that presented the best performances.

As Munaro *et al.* conclude, the need for caution is also highlighted here when considering this factor since there is an indication of a context-dependent nature. When posting a video on YouTube, channel followers are notified about the publication when this option is activated. The hypothesis suggested here is that when this occurs at the end of the week, that is, on Friday or Saturday, the probability of these followers being available to watch the video more readily is greater than during the week.

Finally, regarding the variables linked to the mediator/channel, it is important to highlight a limitation: the existence of a human figure in the thumbnail of the videos does not imply that they are representations or photographs of the channel owners, given that they can also be illustrative figures or be associated with people invited to interact in the production itself.

For this reason, the fact that these figures are mostly male and white explains three possible scenarios in relation to the smaller number of women and non-white people found in the *thumbnails* of the videos analyzed here: a) they were less invited to talk about the vaccine theme as an authority; b) were used less to represent the population/topic discussed in the video, or c) produced less content about vaccines on YouTube. In any case, the structural nature of racism and sexism in Brazilian society is evident concerning the production of videos about vaccines on YouTube (Gonzales, 1984).

In this regard, studies on the implications of gender in science and scientific dissemination have highlighted the existence of asymmetries in both the presence and reception of content/work produced by men and women. Welbourne and Grant (2015), for example, highlight that there is a greater prevalence of men compared to women as authors of science videos on YouTube.

Dalyot, Rozenblum, and Baram-Tsabari (2022), when analyzing the comments made on posts promoting scientific work done by women and men, highlight two factors that may be associated with the lower representation of women on platforms with greater exposure to audiences: the risk of disqualification of work and disadvantageous, inadequate, or hostile reception of female productions.

FINAL CONSIDERATIONS

The present study, by seeking to understand the factors associated with the greater number of views of videos in Brazilian Portuguese, specifically focused on vaccines on YouTube, presents important contributions to the health communication literature.

Initially, this study offers evidence about the potential of using elements linked to disinformation/controversy in video titles to attract audiences based on results from a sample of 13,619 videos, confirming the hypotheses highlighted in previous studies by several authors. Moreover, as previously discussed, science communication channels are using the strategy of adopting elements linked to misinformation/

controversy about vaccines in their titles, which often proves to be an effective means of ensuring a greater number of views for productions.

Furthermore, the present investigation has a practical dimension – its results can provide a basis to better guide the adoption of certain native YouTube variables with the potential to generate a greater number of views, namely: title size, video length, use of characters in capital letters, use of terms associated with science and symptoms, use of human figures in the thumbnail, and ideal day of the week and time of day for publishing videos about vaccines.

However, some limitations need to be highlighted. The first of these concerns the dependent variable, which here was restricted to the number of views, but which needs to be expanded to others, such as the number of likes, dislikes, and comments. Covolo and collaborators (2017), for example, found that, despite being seen less, anti-vaccine videos garnered more likes and shares.

It is also necessary to further investigate the content mobilized in the videos to verify whether what was observed by Johnson and collaborators (2020) on Facebook in relation to the composition of the content produced by pro-vaccine and anti-vaccine groups and its potential to arouse interest or convince undecided people, also applies to YouTube. This could be done through topic modeling or another NLP technique suitable for analyzing large quantities of textual data.

Furthermore, the interpretation regarding race/color and gender ended up being comprehensive when compared to the initial motivation to include them in the investigation. However, these articulations certainly raise important questions about the asymmetries already documented in the literature, produced, and perpetuated by actions based on gender (patriarchy) and race/color (racism). They also highlight the importance of new studies focused on this dimension in health communication, and, more specifically, in productions about vaccines on YouTube.

Thus, the findings of this research corroborate the finding that cyberspace is far from being a neutral or merely technocratic territory. As Rosalía Winocur (2006) points out, the internet is not just translated into virtualities, as it is not closed in itself, it is not hermetically closed to the political, socio-historical, and cultural contexts in which individuals are circumscribed. The internet is produced by "real subjects in the concrete spaces of their everyday lives" (Winocur, 2006, p. 554, our translation). Nor is it a self-referenced culture, but rather one embedded in the practical and symbolic dimensions in which these subjects are situated.

Hence, the importance of an analysis based on a comprehensive sociological reading (Giddens, 1989) is highlighted, which has the premise of understanding what people do with ICT in the face of constraints arising from their functional affordances, bearing in mind that the "logic of the platform" is the expression of the "logic of capital" (Souza, 2020).

REFERENCES

BROTAS, Antonio Marcos Pereira; COSTA, Marcia Cristina Rocha.; MASSARANI, Luisa. Enquadramentos e desinformação sobre vacina contra covid-19 no YouTube: embaralhamentos entre ciência e negacionismo. **Mídia & Cotidiano**, Niterói, v. 15, n. 3, p. 73-100, 2021. DOI: https://doi.org/10.22409/rmc.v15i3.50954. Available from: https://periodicos.uff.br/midiaecotidiano/article/view/50954. Accessed: 15 Dec. 2023.

BUTLER, Judith. **Problemas de gênero**: feminismo e subversão da identidade. 22. ed. Rio de Janeiro: Civilização Brasileira, 2003.

CALVO, Dafne; CANO-ORÓN, Lorena; LLORCA-ABAD, Germán. Covid-19 vaccine disinformation on YouTube: analysis of a viewing network. **Communication & Society**, Pamplona, v. 35, n. 2, p. 223-238, 2022. DOI: https://doi.org/10.15581/003.35.2.223-238. Available from: https://revistas.unav.edu/index.php/communication-and-society/article/view/42099. Accessed: 15 Dec. 2023.

CHAN, Calvin. *et al.* The reliability and quality of YouTube videos as a source of Public Health Information regarding covid-19 vaccination: Cross-sectional study. **JMIR Public Health and Surveillance**, Toronto, v. 7, n. 7, e29942, 2021. DOI: https://doi.org/10.2196/29942. Available from: https://pubmed.ncbi.nlm.nih.gov/34081599/. Accessed: 15 Dec. 2023.

COVOLO, Loredana *et al.* What arguments on vaccinations run through YouTube videos in Italy? A content analysis. **Human Vaccines & Immunotherapeutics**, Austin, v. 13, n. 7, p. 1693-1699, 2017. DOI: https://doi.org/10.1080%2F21645515.2017.1306159. Available from: https://www.tandfonline.com/doi/full/10.1080/21645515.2017.1306159. Accessed: 15 Dec. 2023.

DALPOZ, Gabriel Quirino *et al.* Analysis of YouTube® educational videos on prevention of dental caries. **Research, Society and Development**, v. 11, n. 1, p. e26011124693, 2022. DOI: https://doi.org/10.33448/rsd-v11i1.24693. Available from: https://rsdjournal.org/index.php/rsd/article/view/24693. Accessed: 15 Dec. 2023.

DALYOT, Keren; ROZENBLUM, Yael; BARAM-TSABARI, Ayelet. Engagement patterns with female and male scientists on Facebook. **Public Understanding of Science**, London, v. 31, n. 7, p. 867-884, 2022. DOI: https://doi.org/10.1177/09636625221092696. Available from: https://journals.sagepub.com/doi/full/10.1177/09636625221092696. Accessed: 15 Dec. 2023.

DAMSTRA, Alyt *et al.* What does fake look like? a review of the literature on intentional deception in the news and on social media. **Journalism Studies**, London, v. 22, n. 14, p. 1947-1963, 2021. DOI: https://doi.org/10.1080/1461670X.2021.1979423. Available from: https://www.tandfonline.com/doi/full/10.1080/1461670X.2021.1979423. Accessed: 15 Dec. 2023.

FIALHO, Inês. *et al.* Enfarte agudo do miocárdio no YouTube – Is it all fake news? **Revista Portuguesa de Cardiologia**, v. 40, n. 11, p. 815-825, 1 nov. 2021. Available from: https://www.revportcardiol.org/pt-enfarte-agudo-do-miocardio-no-articulo-S0870255121001542. Accessed: 15 Dec. 2023.

GARBIN, Helena Beatriz da Rocha; PEREIRA NETO, André de Faria; GUILAM, Maria Cristina Rodrigues. A internet, o paciente expert e a prática médica: uma análise bibliográfica. **Interface – Comunicação, Saúde, Educação**, Botucatu, v. 12, n. 26, p. 579-588, 2008. DOI: https://doi.org/10.1590/S1414-32832008000300010. Available from: https://www.scielo.br/j/icse/a/TPC5B5678dnn9YXBFD3KkrK/?lang=pt. Accessed: 15 Dec. 2023.

GIDDENS, Anthony. **A constituição da sociedade**. São Paulo: Martins Fontes, 1989.

GONZALES, Lélia. Racismo e sexismo na cultura brasileira. **Revista Ciências Sociais Hoje**, São Paulo, v. 2, n. 1, p. 223-243, 1984. Available from: https://edisciplinas.usp.br/pluginfile.php/7395422/mod_resource/content/1/GONZALES%2C%20L%C3%A9lia%20-%20Racismo_e_Sexismo_na_Cultura_Brasileira%20%281%29.pdf. Accessed: 15 Dec. 2023.

JOHNSON, Neil F. *et al.* The online competition between pro- and anti-vaccination views. **Nature**, London, v. 582, n. 7811, p. 230-233, 13 May 2020. DOI: https://doi.org/10.1038/s41586-020-2281-1. Available from: https://www.nature.com/articles/s41586-020-2281-1. Accessed: 15 Dec. 2023.

KÄRKKÄINEN, Kimmo; JOO, Junggseock. FairFace: face attribute dataset for balanced race, gender, and age. **arXiv,** Cornell, 14 Aug. 2019. *Preprint.* DOI: https://doi.org/10.48550/arXiv.1908.04913. Available from: https://arxiv.org/abs/1908.04913. Accessed: 15 Dec. 2023.

LI, Heidi Oi-Yee *et al.* YouTube as a source of information on covid-19: a pandemic of misinformation? **BMJ Global Health**, London, v. 5, n. 5, p. e002604, 1 May 2020. DOI: https://doi.org/10.1136/bmjgh-2020-002604. Available from: https://pubmed.ncbi.nlm.nih.gov/32409327/. Accessed: 15 Dec. 2023.

LIMA, Clóvis Ricardo Montenegro de *et al.* Emergência de saúde pública global por pandemia de covid-19. **Folha de Rosto: Revista de Biblioteconomia e Ciência da Informação**, Juazeiro do Norte, v. 6, n. 2, p. 5-21, 2020. DOI: https://doi.org/10.46902/2020n2p5-21. Available from: https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/490. Accessed: 15 Dec. 2023.

LOPES, Arthur da Silva; BROTAS, Antonio Marcos Pereira. Echo chambers and vaccines against covid-19 mis/disinformation on Twitter: machine learning and network analysis-based approach. **Research, Society and Development**, Vargem Grande Paulista, v. 12, n. 2, p. e22812240159, 2023. DOI: https://doi.org/10.33448/rsd-v12i2.40159. Available from: https://rsdjournal.org/index.php/rsd/article/view/40159. Accessed: 15 Dec. 2023.

MASSARANI, Luisa; COSTA, Márcia Cristina Rocha.; BROTAS, Antonio Marcos Pereira. A pandemia de covid-19 no YouTube: ciência, entretenimento e negacionismo. **Revista Latinoamericana de Ciencias de la Comunicación**, São Paulo, v. 19, n. 35, p. 2455-256, 2020. Available from: https://www.arca.fiocruz.br/handle/icict/46507. Accessed: 15 Dec. 2023.

MASSARANI, Luisa *et al.* Vacinas contra a covid-19 e o combate à desinformação na cobertura da Folha de S.Paulo. **Fronteiras – Estudos Midiáticos**, São Leopoldo, v. 23, n. 2, p. 29-43, 2021. DOI: https://doi.org/10.4013/fem.2021.232.03. Available from: https://revistas.unisinos.br/index.php/fronteiras/article/view/22592. Accessed: 15 Dec. 2023.

MCCLUNG, H. J.; MURRAY, R. D.; HEITLINGER, L. A. The Internet as a source for current patient information. **Pediatrics**, Elk Grove Village, v. 101, n. 6, jun. 1998. DOI: https://doi.org/10.1542/peds.101.6.e2. Available from: https://pubmed.ncbi.nlm.nih.gov/9606244/. Accessed: 15 Dec. 2023.

MUNARO, Ana Cristina. *et al.* To engage or not engage? The features of video content on YouTube affecting digital consumer engagement. **Journal of Consumer Behaviour**, London, v. 20, n. 5, p. 1336-1352, 2021. DOI: https://doi.org/10.1002/cb.1939. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cb.1939. Accessed: 15 Dec. 2023.

NEWMAN, Nic. Overview and key findings of the 2022 – Digital News Report. **Reuters Institute for the Study of Journalism,** London, 15 jun. 2022. Available from: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022/dnr-executive-summary. Accessed: 23 nov. 2022.

NOGUEIRA, Oracy. Preconceito racial de marca e preconceito racial de origem: sugestão de um quadro de referência para a interpretação do material sobre relações raciais no Brasil. **Tempo Social**, São Paulo, v. 19, n. 1, p. 287-308, 2007. DOI: https://doi.org/10.1590/S0103-20702007000100015. Available from: https://www.scielo.br/j/ts/a/MyPMV9Qph3VrbSNDGvW9PKc/?lang=pt. Accessed: 15 Dec. 2023.

NUNES, Georgina Helena Lima. Autodeclarações e comissões: responsabilidade procedimental dos/as gestores/as de ações afirmativas. *In*: DIAS, Gleidson Renato Martins; TAVARES JUNIOR, Paulo Roberto Faber (org.). **Heteroidentificação e cotas raciais**: dúvidas, metodologias e procedimentos. Canoas: IFRS *Campus* Canoas, 2018. p. 11-30. Available from: https://www.geledes.org.br/wp-content/uploads/2019/03/Heteroidentificacao_livro_ed1-2018.pdf. Accessed: 15 Dec. 2023.

PENTEADO, Claudio Luis de Camargo *et al*. #Vacinar ou não, eis a questão! As emoções na disputa discursiva sobre a aprovação das vacinas contra a covid-19 no Twitter. **Política & Sociedade**, Florianópolis, v. 20, n. 49, p. 104-133, 2021. DOI: https://doi.org/10.5007/2175-7984.2021.85145. Available from: https://periodicos.ufsc.br/index.php/politica/article/view/85145. Accessed: 15 Dec. 2023.

PEREIRA NETO, André *et al.* O paciente informado e os saberes médicos: um estudo de etnografia virtual em comunidades de doentes no Facebook. **História, Ciências, Saúde – Manguinhos**, Rio de Janeiro, v. 22, supl., p. 1653-1671, 2015. DOI: https://doi.org/10.1590/S0104-59702015000500007. Available from: https://www.scielo.br/j/hcsm/a/NMrcHvYypNG3sFQmvYwv4vR/?format=pdf&lang=pt. Accessed: 15 Dec. 2023.

PIERRI, Francesco. *et al.* Online misinformation is linked to early covid-19 vaccination hesitancy and refusal. **Scientific Reports**, London, v. 12, n. 5966, p. 1-7, 26 abr. 2022. DOI: https://doi.org/10.1038/s41598-022-10070-w. Available from: https://www.nature.com/articles/s41598-022-10070-w. Accessed: 15 Dec. 2023.

RECUERO, Raquel *et al.* **Desinformação, mídia social e covid-19 no Brasil**: relatório, resultados e estratégias de combate. Pelotas: Midiars – Grupo de Pesquisa em Mídia, Discurso e Análise de Redes Sociais, 2021. Relatório de pesquisa. Available from: https://wp.ufpel.edu.br/midiars/files/2021/05/Desinformac%CC%A7a%CC%83o-covid-midiars-2021-1.pdf. Accessed: 15 Dec. 2023.

RECUERO, Raquel; SOARES, Felipe. O discurso desinformativo sobre a cura da covid-19 no Twitter: estudo de caso. **E-Compós**, Brasília, DF, v. 24, p. 1-29, 2021. DOI: https://doi.org/10.30962/ec.2127. Available from: https://www.e-compos.org.br/e-compos/article/view/2127. Accessed: 15 Dec. 2023.

RIEDER, Bernhard. **YouTube Data Tools**. Amsterdam: Digital Methods Initiative, c2023. Available from: https://ytdt.digitalmethods.net/. Accessed: 31 maio 2023.

SERENGIL, Sefik Ilkin; OZPINAR, Alper. LightFace: A hybrid deep face recognition framework. *In*: INNOVATIONS IN INTELLIGENT SYSTEMS AND APPLICATIONS CONFERENCE, 15-17 out. 2020, Istanbul. **Proceedings** [...], Istanbul, v. 2020, p. 1-5, 2020. DOI: http://dx.doi.org/10.1109/ASYU50717.2020.9259802. Available from: https://ieeexplore.ieee.org/document/9259802. Accessed: 15 Dec. 2023.

SOUZA, Rafael Bellan Rodrigues de. A comunicação contra-hegemônica no capitalismo digital: limites e contradições. **Liinc em Revista**, Rio de Janeiro, v. 16, n. 1, p. e5133, 2020. DOI: https://doi.org/10.18617/liinc.v16i1.5133. Available from: https://revista.ibict.br/liinc/article/view/5133. Accessed: 15 Dec. 2023.

VAN DER LINDEN, Sander. Misinformation: susceptibility, spread, and interventions to immunize the public. **Nature Medicine**, London, v. 28, n. 3, p. 460-467, 2022. DOI: https://doi.org/10.1038/s41591-022-01713-6. Available from: https://www.nature.com/articles/s41591-022-01713-6. Accessed: 15 Dec. 2023.

WARDLE, Claire; DERAKHSHAN, Hossein. **Information disorder**: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe, 2017. Available from: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html. Accessed: 31 maio 2023.

WELBOURNE, Dustin J.; GRANT, Will J. Science communication on YouTube: Factors that affect channel and video popularity. **Public Understanding of Science**, London, v. 25, n. 6, p. 706-718, 2015. DOI: http://dx.doi.org/10.1177/0963662515572068. Available from: https://journals.sagepub.com/doi/10.1177/0963662515572068. Accessed: 15 Dec. 2023.

WORLD HEATH ORGANIZATION (WHO). **Ten threats to global health in 2019**. Geneva: WHO, 2019. Available from: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019. Accessed: 23 nov. 2022.

WINOCUR, Rosalía. Internet en la vida cotidiana de los jóvenes. **Revista Mexicana de Sociología**, Mexico City, v. 68, n. 3, p. 551-580, 2006. Available from: https://www.redalyc.org/pdf/321/32112601005.pdf. Accessed: 15 Dec. 2023.

YANG, Shiyu *et al.* The science of YouTube: What factors influence user engagement with online science videos? **PLoS ONE**, San Francisco, v. 17, n. 5, p. e0267697, 2022. DOI: https://doi.org/10.1371/journal.pone.0267697. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0267697. Accessed: 15 Dec. 2023.