VISUAL REVIEW | Vol. 17, No. 3, 2025 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v17.5786

CENSORSHIP AND SELF-CENSORSHIP OF UNIVERSITY STUDENTS ON INSTAGRAM AND TIKTOK

LUIS ANTONIO LÓPEZ-FRAILE [1], MIGUEL ÁNGEL ALONSO GUISANDE [1]
[1]European University of Madrid, Spain

KEYWORDS

Censorship
Self-censorship
Social media
Digital behaviour

ABSTRACT

This study investigates the censorship and self-censorship of young university students on two of the most popular social media platforms: Instagram and TikTok. The research focused on how these students perceive censorship on both platforms and how this perception translates into self-censorship behaviour. It explores the perceived similarities and differences between the two networks and determines which one presents a more restrictive environment for freedom of expression. A non-experimental study was conducted with a sample of 502 students. The findings reveal a unanimous perception of censorship and a corresponding self-censorship response among the young participants.

Received: 11/ 02 / 2025
Accepted: 13/ 05 / 2025

1. Introduction

The objective of this research is to analyse the phenomenon of censorship and self-censorship among young university students on the social networks Instagram and TikTok. The study will examine the content control practices implemented by the platforms, as well as young people's perception of censorship, self-censorship and the influence of both on their online behaviour.

The issue of internet control and censorship has become a matter of growing global concern, particularly as most internet users are young people (Vizcaíno-Laorga et al., 2019). A diverse array of entities and governments have adopted regulatory measures to control online content, impacting freedom of expression, access to information, and privacy. Social networks, such as Instagram and TikTok, employ artificial intelligence systems for content moderation, giving rise to ethical and legal challenges related to censorship (Conde, 2024). In this paper, an attempt has been made to assist the scientific community in clarifying the issue by providing an analysis of data collected from young university students in the Madrid region through a survey on their online behaviour on both social networks.

1.1. Censorship on Social Media

Social media platforms, as private entities, have the capacity to establish their own rules of use within the framework of the contracts of adhesion they conclude with users. While this could be defended on the basis of the autonomy of private will, it is also logical to defend that this autonomy is subject to legality, morality and public order. The content control exercised by private entities on social networks has been shown to have effects comparable to state censorship, even amplified by the viral nature of the Internet (MacKinnon, 2012). In such cases, these private entities should assume responsibility for content uploaded or created by users when it has not been blocked or deleted, as they have control over it and have exercised active moderation (San Juan, 2021).

In light of these considerations, it is crucial to establish limits to the autonomy of the private will in order to protect fundamental rights, especially since the Internet has become the main source of access to information, so that the right to private autonomy, on the one hand, and freedom of expression, recognised in all Western constitutions, on the other, seem to be in conflict. The automatic filtering of content, although justified by the fight against fake news, piracy or child pornography, poses a significant risk to the fundamental rights of users, especially freedom of expression and information, since "without access to information, genuine freedom of expression is not possible" (Sturges, 2010, p. 21).

Some authors, such as San Juan (2021), address the evolution of censorship in the context of digital platforms. The author posits that, while traditional censorship has historically been associated with the intervention of state and religious powers, contemporary social networks wield significant control over the dissemination of information, a development that gives rise to grave concerns regarding freedom of expression and access to information. The prohibition of prior censorship, enshrined in Article 20.2 of the Spanish Constitution, does not apply in the same way to the actions of private companies that manage social networks. These platforms, by filtering and blocking content, can have comparable and even more severe effects than state censorship, due to the expansive nature of the Internet and its capacity to amplify the dissemination of information. Furthermore, the discourse encompasses the phenomenon of "fake news" and its repercussions on freedom of expression. Whilst content control may be intended to combat disinformation, it also carries the risk of violating fundamental rights. The author proposes a balanced approach, integrating content regulation with the promotion of education and media literacy among users (San Juan, 2021).

In addition, other authors examine how content moderation decisions on social media are influenced by cultural values and regulatory pressures (Gillespie, 2018). Balkin (2004) proposes a theory of freedom of expression in the information society, exploring how to preserve it in the digital context. Furthermore, analysis of the social networking strategies employed by audiovisual platforms is undertaken, with a particular emphasis on the idiosyncrasies of each platform (Martínez-Sánchez et al, 2021). Other authors draw attention to the distinction between censorship and the Streisand effect, a paradoxical phenomenon whereby attempts to censor information can result in its wider dissemination, thus marking a milestone in the understanding of perceived censorship (Stewart and Bunton, 2016).

The role of Artificial Intelligence in content moderation is crucial. Algorithms are utilised to detect and remove content that contravenes platform rules and is deemed inappropriate, including hate speech, violence, child pornography, and misinformation (Jansen and Martin, 2015). AI systems operate through two primary mechanisms: matching and classification. Matching involves the comparison of content with a database of examples, while classification analyses patterns in the data to identify problematic content without the need for an exact match. Classification is the most widely used technique in social media content moderation (Gorwa et al., 2020). The utilisation of AI in this domain offers several advantages, including cost reduction, enhanced content coverage, optimised efficiency, and expedited detection of inappropriate content (Llansó et al., 2020). While this can be useful in combating illegal content, it can also lead to censorship of legitimate expression (Rosales et al., 2024).

1.2. Self-Censorship on Social Media

Self-censorship is a complex phenomenon that occurs when users limit their own expression for fear of negative consequences. This fear can come from different sources, such as the possibility of being criticised, attacked or even blocked by the platform (Pérez et al., 2019).

Based on the sources consulted, we shall take a look at some key points about self-censorship on social networks. For instance, users may self-censor due to a fear of negative reactions, judgement, criticism, or even attack for their opinions. Furthermore, self-censorship can emerge as a means to avoid conflicts with family, friends or at work, as well as to avoid damaging one's reputation (Gomes-Franco-Silva and Sendín-Gutiérrez, 2014). Some authors argue that today's culture promotes inclusion, which can lead people to censor themselves in order to avoid social exclusion. The pressure to adhere to political correctness can result in the avoidance of sensitive issues or the modification of language to avoid offending others. Self-censorship can be seen as a filter that people impose on themselves to avoid problems, as social networks allow users to read their own posts before they are published, which can lead to self-regulation or self-censorship. This filter can be internal, in which case the individual determines what not to publish, or it can be imposed by society. In the online sphere, users wield greater autonomy over their expression, yet they are also more vulnerable to public scrutiny (Serrano, 2016).

This phenomenon prompts the consideration that self-censorship can impede freedom of expression, prompting individuals to refrain from articulating their ideas due to a sense of fear. While some authors advocate for the importance of uninhibited expression, they also emphasise the need for people to be aware of the impact their words may have on others, and to modulate the tone of their arguments so as not to offend. Conversely, others advocate for an unbridled freedom of expression, devoid of any form of moderation, provided it does not transgress legal boundaries (Pérez and El Mecky, 2024).

In summary, self-censorship on social networks is a complex phenomenon that reflects the tension between the need to express oneself freely and the fear of social consequences. While self-censorship can be a means of avoiding conflict, it can also limit the diversity of opinions and freedom of expression online.

2. Methodology and Objectives

The fundamental objective of this study is to ascertain how young university students perceive the censorship generated by the TikTok and Instagram platforms and how this perception manifests as self-censorship behaviour in response. The central aim is to ascertain the similarities and differences perceived by young people between the two networks and which of them offers a more or less restrictive environment in terms of the combination of freedom of expression and respectful behaviour.

To achieve the proposed objective, after conducting a descriptive theoretical analysis at the beginning of this research, we carried out fieldwork consisting of analysing the perception of censorship and self-censorship, measured in key aspects such as whether they have been subject to censorship in their publications on both social networks, in which one they believe censorship is more intense, whether they perceive it to be preceding, simultaneously or a posteriori to the publication of posts, whether they have modified their behaviour when generating content based on previous experiences of censorship, whether they perceive the censorship activity of both platforms as fair or unfair, whether they believe that such censorship is in pursuit of laudable or spurious objectives, and whether they resort to trickery to circumvent the platforms' possible censorship activity.

The fieldwork for this study was conducted through a non-experimental research design, utilising an online questionnaire with four types of alternative responses based on the Likert scale (A LOT/MODERATELY/LITTLE/NONE). The study sample was drawn from a convenience sample of 502 students from various universities in the Community of Madrid. The questionnaire comprised a series of 24 questions (listed below from 1 to 24). This methodology will allow the researchers to obtain numerical data on the perception of censorship and self-censorship among young university students, as well as to carry out statistical analyses to identify patterns and relationships between the variables studied. While the primary focus of our research is on the experience of young university students in the Community of Madrid, the results can be generalised, with a high degree of reliability, to other student populations, given the substantial number of questionnaires collected, as well as the affiliation of the respondents, who belonged to both public and private universities throughout the region.

However, it is important to note that the fieldwork was conducted in accordance with the ethical principles of human research. Prior to the administration of the survey, written informed consent was obtained from all participants, and the anonymity and confidentiality of the data collected was guaranteed.

The survey was meticulously designed to collect pertinent information on:

- Perceptions of censorship: students were asked about their awareness of Instagram and TikTok's content moderation policies, and whether they believed these policies were applied fairly and transparently.
- Experiences of censorship: we asked whether students had experienced censorship on these platforms, either through content removal, account suspension or limited visibility.
- Self-censorship: students were invited to indicate how frequently they self-censor on Instagram and TikTok, and to specify the primary motivations behind such self-censorship.
- Elements of self-censorship: Our study explored variables related to gender, political ideology, xenophobia, the human body and violent communication as drivers of self-censorship.

In a subsequent section of this paper, the results obtained from the field research will be presented and analysed, providing the indicative conclusions drawn from the answers given by the respondents.

The questionnaire was initiated with a comprehensive, academic, and scientific definition of the terms "censorship" and "self-censorship" in the context of social media. This definition was provided for the purposes of this study, ensuring that respondents could associate the issues raised with their own practices concerning the consumption and creation of content on the two social networks under study. Prior to the commencement of the survey, respondents were informed that the term "censorship" in the context of social networks is defined as "the suppression or modification of communicative content (texts, images, speeches, etc.) that is considered offensive, subversive, politically unacceptable or harmful to the public good", as well as "the control of content and the blocking of user accounts that is carried out in social networks" (San Juan, 2021, p. 23).

Self-censorship in social networks can be defined as "the voluntary restriction that an individual or institution exercises over its own expressions for fear of social, legal or economic reprisals" (Pino et al, 2022, p. 128).

In the survey's introduction, the terms "gender," "political ideology," "xenophobia," "human body," and "violent communication" were defined for the study's participants, ensuring a precise interpretation of the research questions. Specifically, it was explained that:

- The gender-related question aims to investigate instances of self-censorship pertaining to gender ideology and behaviour, or attitudes that may be ethically or punitively reprehensible in relation to gender.
- The question on political ideology aims to investigate instances of self-censorship in aspects related to ideological criticism, identification of the respondent's political ideology or political stance.
- The question on xenophobia aims to investigate self-censoring behaviour in any manifestation bordering on offensive behaviour towards other people, linked to race, religion or geographical origin, in any environment (social, sporting, artistic, cinematographic, etc.).

- The question on the human body aims to investigate self-censorship in relation to the representation of the body or parts of the body, understanding them solely in terms of their sexual function.
- The question on violent communication aims to investigate self-censorship in relation to the use of offensive, hurtful, hateful or violence-inciting language.

The questionnaire comprised the following questions, designed to address all aspects of this research, with the aim of enabling us to draw reliable conclusions that respond to the stated objectives:

1. To what extent are you aware of the content moderation policies on Instagram and TikTok?
2. To what extent do you believe these policies are fair, transparent, and well-explained?
3. To what extent do you experience censorship on Instagram, whether through content removal, account suspension, or reduced visibility?
4. To what extent do you experience censorship on TikTok, whether through content removal, account suspension, or reduced visibility?
5. To what extent do you experience censorship on the Instagram platform before posting content or writing a comment on a post?
6. To what extent do you experience censorship on the TikTok platform before posting content or writing a comment on a post?
7. To what extent do you experience censorship on the Instagram platform while posting content or writing a comment on a post?
8. To what extent do you experience censorship on the TikTok platform while posting content or writing a comment on a post?
9. To what extent do you experience censorship on the Instagram platform after posting content or writing a comment on a post?
10. To what extent do you experience censorship on the TikTok platform after posting content or writing a comment on a post?
11. To what extent have you modified your behaviour when creating content based on prior experiences of censorship on Instagram?
12. To what extent have you modified your behaviour when creating content based on prior experiences of censorship on TikTok?
13. To what extent do you self-censor due to gender-related issues on Instagram?
14. To what extent do you self-censor due to gender-related issues on TikTok?
15. To what extent do you self-censor due to political ideology issues on Instagram?
16. To what extent do you self-censor due to political ideology issues on TikTok?
17. To what extent do you self-censor due to xenophobia-related issues on Instagram?
18. To what extent do you self-censor due to xenophobia-related issues on TikTok?
19. To what extent do you self-censor your body on Instagram?
20. To what extent do you self-censor your body on TikTok?
21. To what extent do you self-censor violent communication on Instagram?
22. To what extent do you self-censor violent communication on TikTok?
23. To what extent do you attempt to evade the censorship activities of the Instagram platform?
24. To what extent do you attempt to evade the censorship activities of the TikTok platform?

In the subsequent section, the results obtained will be subjected to detailed analysis and interpretation, with a view to drawing conclusions that are commensurate with the object of study in our research.

3. Data Analysis and Results

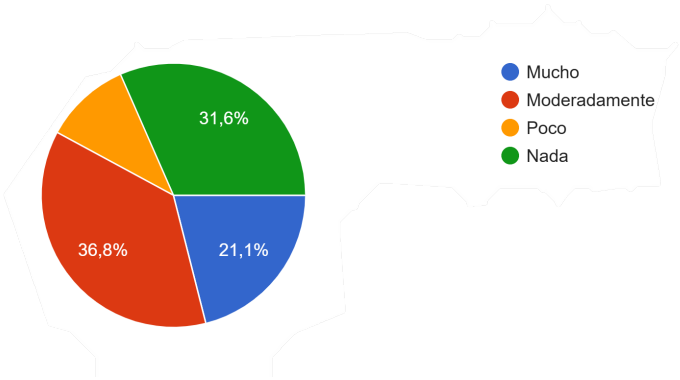
This section presents the results obtained during the fieldwork in the data collection stage, in which an online questionnaire was designed using the Google Forms tool.

The questionnaires were disseminated to the students selected as the study population from 2 to 20 December 2024, resulting in a total of 505 completed questionnaires. Following the processes of data

organisation, data cleansing, and the elimination of questionnaires with unanswered responses, a total of 502 valid questionnaires remained. The subsequent presentation of the items contained in the questionnaire is accompanied by the implementation of graphs as the optimal visualisation element, on which axiomatic analyses are carried out according to the results represented in each one of them.

The questionnaire commences with an initial question regarding the respondent's familiarity with the moderation policies for content disseminated on Instagram and TikTok social media platforms. This serves to establish the respondents' awareness of these regulations and the importance they attribute to them. As illustrated in Figure 1, while the results demonstrate a certain degree of equality, a higher percentage of respondents report a high or moderate degree of knowledge about the issue, with a total of 57.9%, compared to those who report having no or little interest, with a total of 42.1%, in the policies for publishing content on Instagram and TikTok.

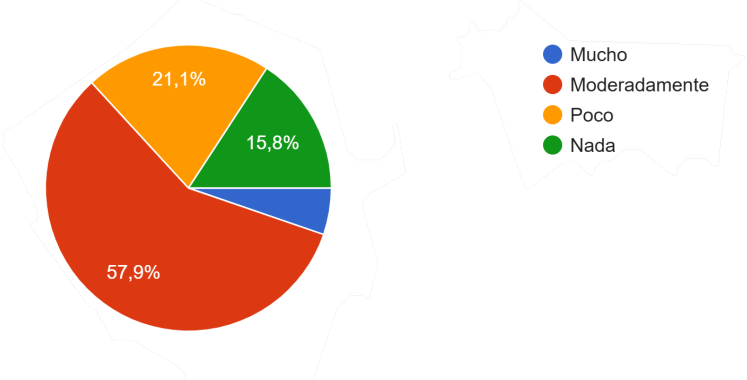
Figure 1. To what extent are you aware of the content moderation policies on Instagram and TikTok?



Source: Own elaboration, 2025.

The second question, illustrated in figure 2, enquires as to the extent to which respondents consider the content publication policies on Instagram and TikTok to be fair and transparent. A mere 5.3% of respondents consider them to be fair and well explained, however, when combined with the 57.9% of respondents who consider them to be moderately fair, this results in a total of 63.2%. This suggests that, in general, young people consider them to be acceptable. The percentage of 36.9% who think the opposite is not negligible, so we could ask the platforms to improve this aspect by increasing transparency.

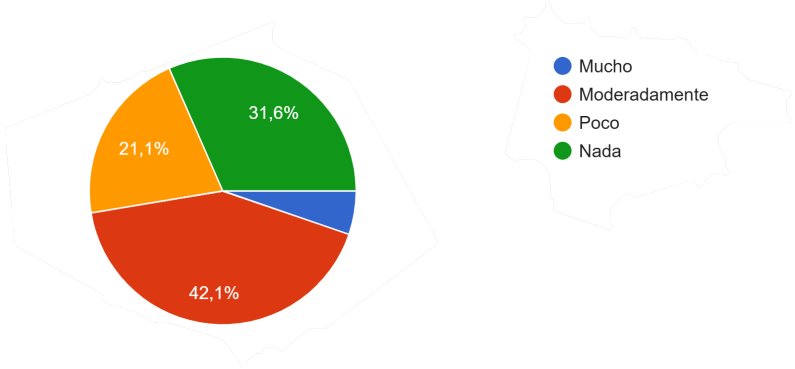
Figure 2. To what extent do you believe these policies are fair, transparent, and well-explained?



Source: Own elaboration, 2025.

In question 3, respondents are invited to provide reflections on their experiences of censorship on Instagram. The results indicate that 47.4% of respondents have encountered some form of restriction on their publications. This finding suggests that the limitation of content to users in various professional domains is a prevalent practice on this social network and constitutes a component of the community standards delineated on the platform itself (Instagram, 2025).

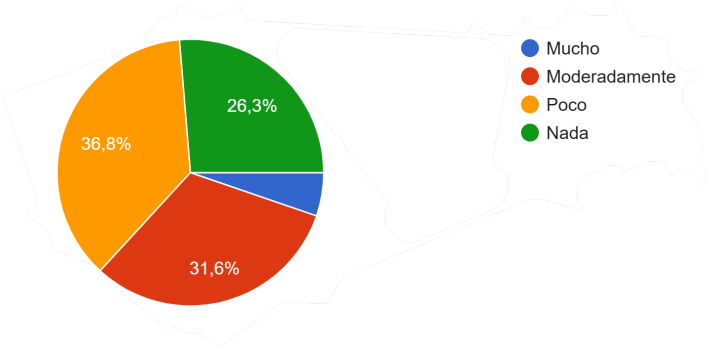
Figure 3. To what extent do you experience censorship on Instagram, whether through content removal, account suspension, or reduced visibility?



Source: Own elaboration, 2025.

In this particular line of enquiry, the fourth question pertaining to the social network TikTok reveals that 36.9% of respondents have observed limitations in their interactions on this social network. However, this figure is notably lower than the proportion cited above for the Instagram social network. The analysis indicates that the predominant experience of censorship across both networks falls within the middle range, categorised as "Little" or "Moderately," with a marginal number perceiving censorship to be high.

Figure 4. To what extent do you experience censorship on TikTok, whether through content removal, account suspension, or reduced visibility?

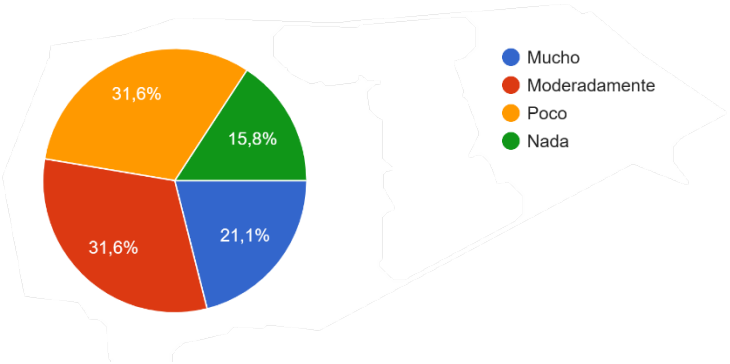


Source: Own elaboration, 2025.

In questions 5, 6, 7, 8, 9 and 10, the questionnaire design is implemented to obtain data from the study population regarding the time variable: before the publication, at the time of publication and after publication in both social networks under study. With regard to the previous time instant, the subjects' feeling of censorship is greater on Instagram with 51.7% compared to 26.3% on TikTok (see figures 5 and 6), a significant difference, which contrasts with the data shown in figures 7 and 8, which refer to the time instant while the action of adding content to the social network is executed. In this case, the percentages are almost identical on both social networks, with 63.1% of the subjects surveyed.

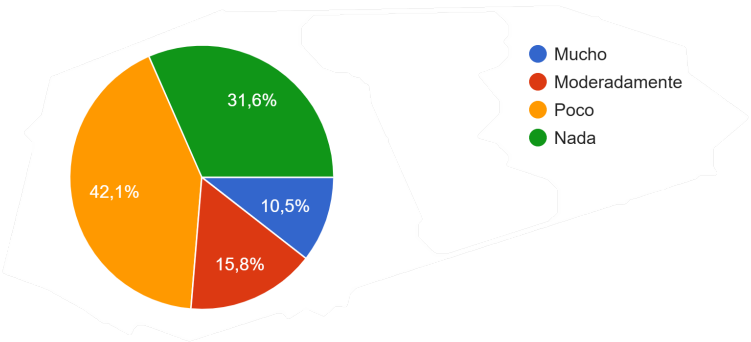
The final term, pertaining to the time variable of analysis, referring to the instant after the publication, exhibited analogous values to those of "A lot" and "Moderately" across both social networks. However, a discrepancy was observed in Instagram, where respondents reported a total of 57.9%, in contrast to the 52.6% recorded in TikTok. Overall, the study population demonstrated a heightened perception of censorship and limitations imposed on Instagram compared to TikTok in the time windows before and after posting. However, the user experience evinces a similar sentiment of censorship when posting on both social networks.

Figure 5. To what extent do you experience censorship on the Instagram platform before posting content or writing a comment on a post?



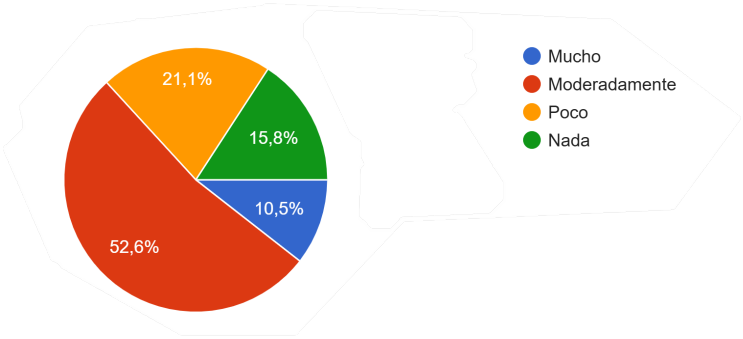
Source: Own elaboration, 2025.

Figure 6. To what extent do you experience censorship on the TikTok platform before posting content or writing a comment on a post?



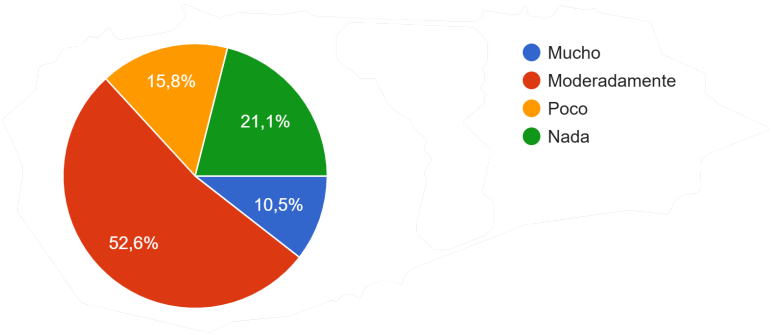
Source: Own elaboration, 2025.

Figure 7. To what extent do you experience censorship on the Instagram platform while embedding content or commenting on a post?



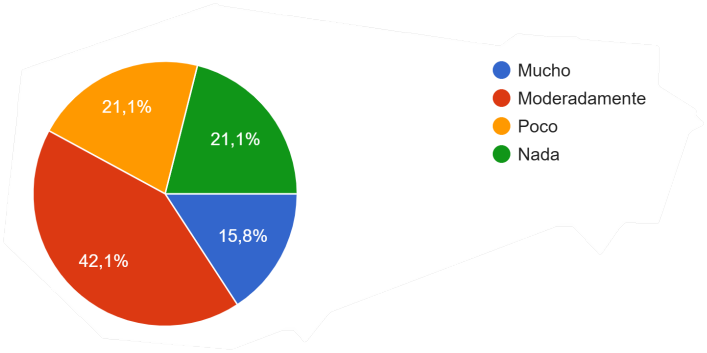
Source: Own elaboration, 2025.

Figure 8. To what extent do you experience censorship on the TikTok platform while posting content or writing a comment on a post?



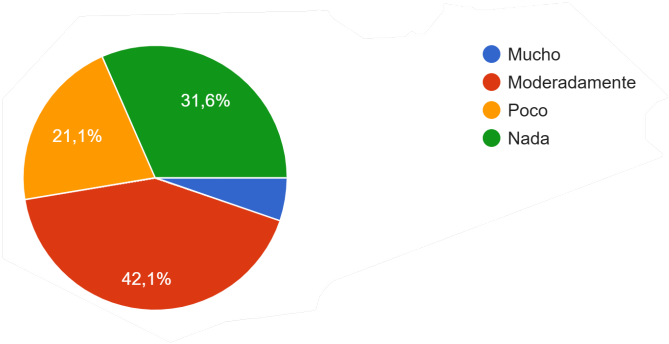
Source: Own elaboration, 2025.

Figure 9. To what extent do you experience censorship on the Instagram platform after posting content or writing a comment on a post?



Source: Own elaboration, 2025.

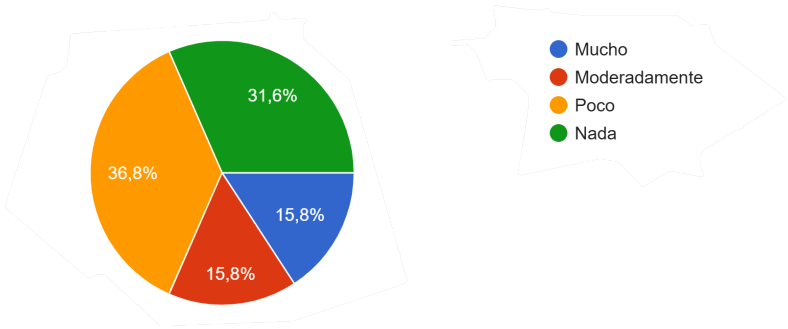
Figure 10. To what extent do you experience censorship on the TikTok platform after posting content or writing a comment on a post?



Source: Own elaboration, 2025.

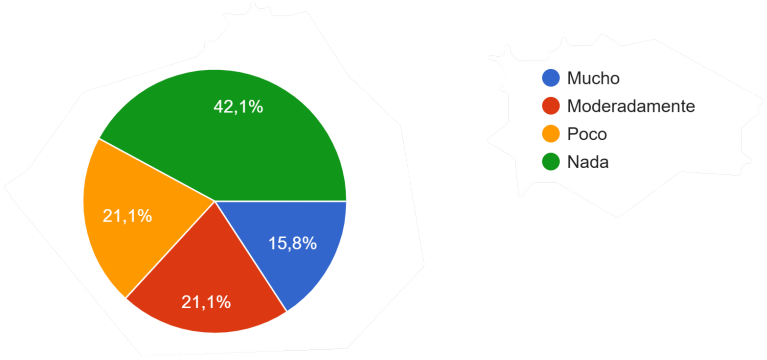
In questions 11 and 12 of the questionnaire, respondents were invited to report on behaviour on both social networks based on their previous experiences (see figures 11 and 12). It is noteworthy that 42.1% of respondents indicated that they do not condition their interaction on the TikTok social network on the basis of previous experiences. This item represents a change in trend and a turning point in the questionnaire, in which, until this question, one could sense a greater inclination, based on the results obtained and the references consulted in other authors, such as Heath (Heath, 2022), towards less intrusion by the content moderation algorithm on the TikTok platform.

Figure 11. To what extent have you modified your behaviour when creating content based on prior experiences of censorship on Instagram?



Source: Own elaboration, 2025.

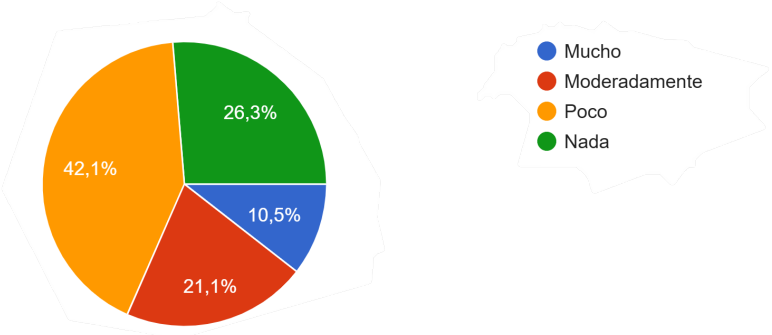
Figure 12. To what extent have you modified your behaviour when creating content based on prior experiences of censorship on TikTok?



Source: Own elaboration, 2025.

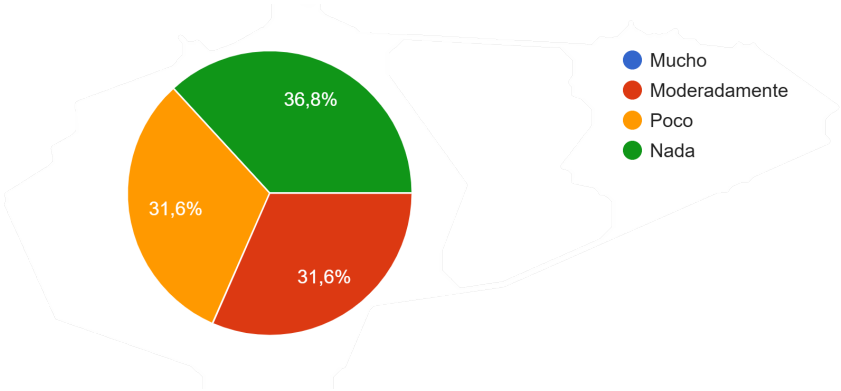
Questions 13 and 14 address the subjects of gender and self-censorship. It is evident that social networks often function as digital spaces that reproduce and amplify issues of gender inequality. Particularly noteworthy is the perception of Instagram as a hostile environment for women, characterised by a notable pressure to adhere to hegemonic patterns and beauty standards (Piñeiro-Otero & Martínez-Rolán, 2024). The data obtained demonstrate that there is a state of equality in the social networks under study, since in both social networks, the values of "A lot" and "Moderately" obtain the same percentage, with a total sum of 31.6% of respondents.

Figure 13. To what extent do you self-censor due to gender-related issues on Instagram?



Source: Own elaboration, 2025.

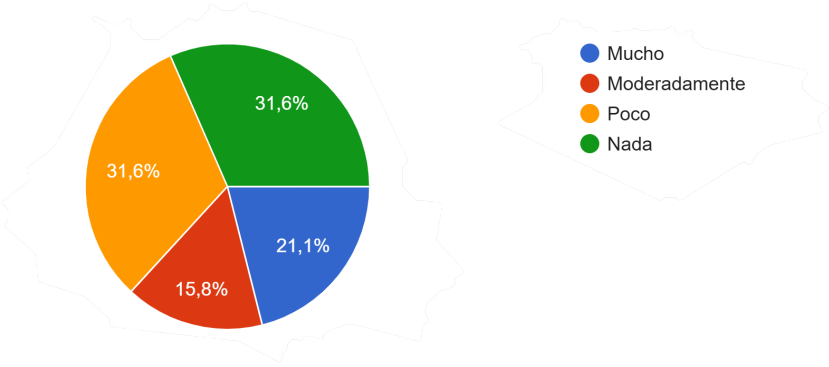
Figure 14. To what extent do you self-censor due to gender-related issues on TikTok?



Source: Own elaboration, 2025.

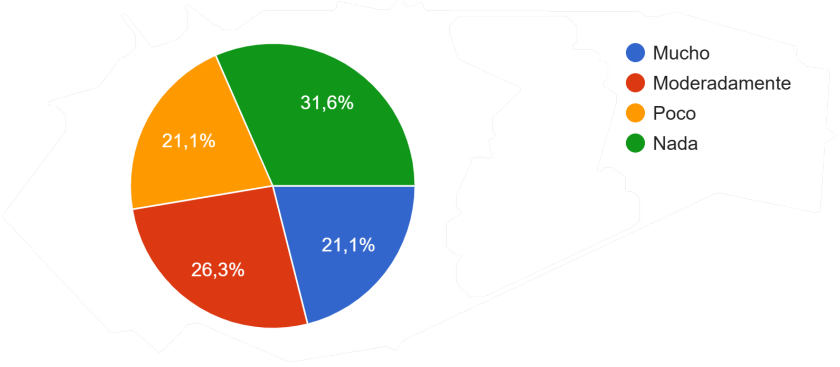
Questions 15 and 16 address the issue of political self-censorship. The phenomenon of political self-censorship in social media is an intrinsic feature of these networks, caused by hate messages that undermine freedom of expression, which should be a goal in all democratic societies (Martínez-Valerio and Mayagoitia-Soria, 2021). In a political context, Instagram and TikTok have been shown to be potent instruments for the dissemination of political information among young people, with the potential to induce self-censorship among ordinary users in the interest of aligning with relevant figures, leaders, and political influencers (Bilewicz and Soral, 2020). The present study reveals a higher percentage of respondents, 10.5%, expressing moderate self-censorship on the social network TikTok compared to the data obtained on Instagram.

Figure 15. To what extent do you self-censor due to political ideology issues on Instagram?



Source: Own elaboration, 2025.

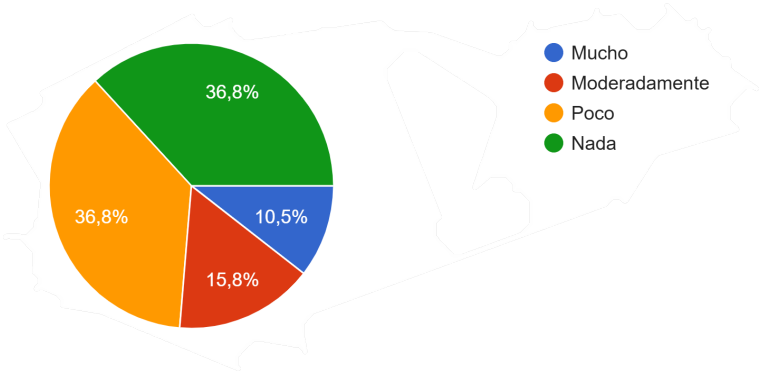
Figure 16. To what extent do you self-censor due to political ideology issues on TikTok?



Source: Own elaboration, 2025.

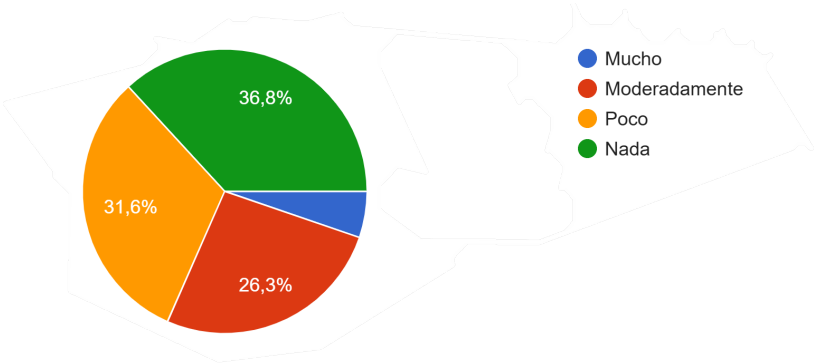
It is important to note that items 17 and 18, with the figures shown in the corresponding figures, are related to the data previously presented in this paper and analysed in figures 15 and 16. Previous studies and academic research have demonstrated the relationship between xenophobia and political polarisation with hate speech on social networks (Evolvi, 2019). This assertion is corroborated by the findings presented in Figures 17 and 18, which illustrate comparative data for the "Moderately" category on Instagram and TikTok. In comparison, the data obtained for the "A lot" value shows a lower net percentage of 10.6% of respondents reporting less self-censorship for xenophobic issues on both social networks.

Figure 17. To what extent do you self-censor due to xenophobia-related issues on Instagram?



Source: Own elaboration, 2025.

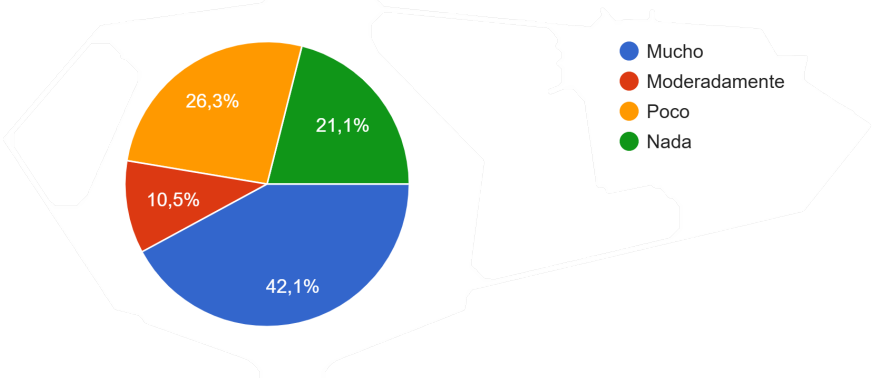
Figure 18. To what extent do you self-censor due to xenophobia-related issues on TikTok?



Source: Own elaboration, 2025.

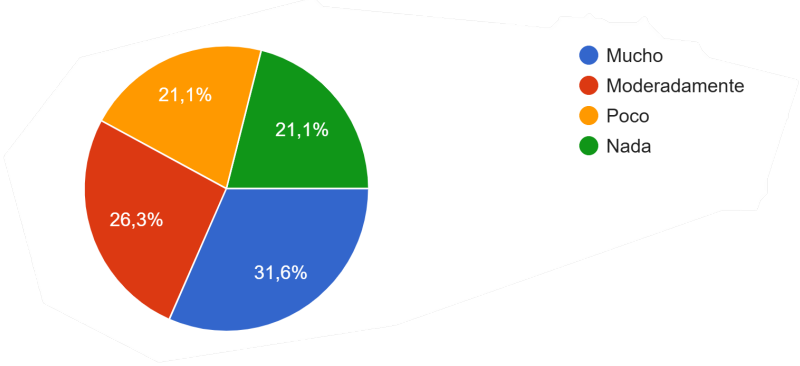
Instagram is a social network that has sought to differentiate itself from other social networks since its inception. It has done so by clearly identifying itself as a platform that prioritises images, in a neoliberal social context where individualism and the generous politicisation of the male and female body prevail (Bard and Magallanes, 2021). Paradoxically, it is a social network with a high degree of intervention in the publication of content if the images explicitly show certain areas of the male or female body, showing a sexualised bias (Sánchez-Holgado and Benito, 2024). This assertion is substantiated by the data presented in Figure 19, which indicates that 42.1% of respondents self-censor their content on Instagram in relation to issues concerning the body. It is also noteworthy that data relating to the social network TikTok shows that 31.6% of students surveyed self-censor their bodies to a high degree, which, when combined with the 26.3% of students who indicate that they moderate their content moderately, exceeds the total percentage of Instagram users who self-censor. This would allow us to define the social network TikTok as a platform that incorporates a high rate of intervention in matters relating to the display of content in which the body or parts of it are shown in publications.

Figure 19. To what extent do you self-censor your body on Instagram?



Source: Own elaboration, 2025.

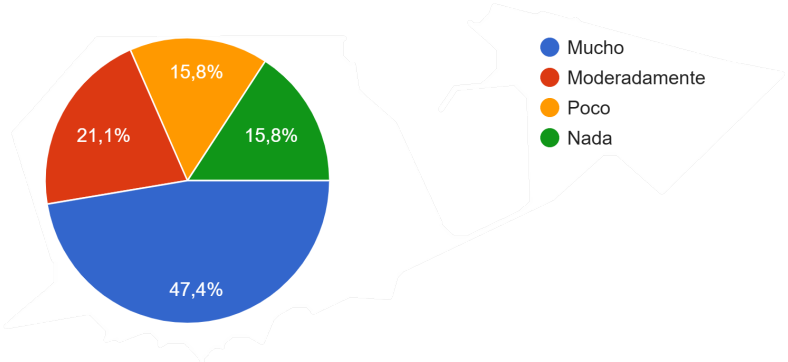
Figure 20. To what extent do you self-censor your body on TikTok?



Source: Own elaboration, 2025.

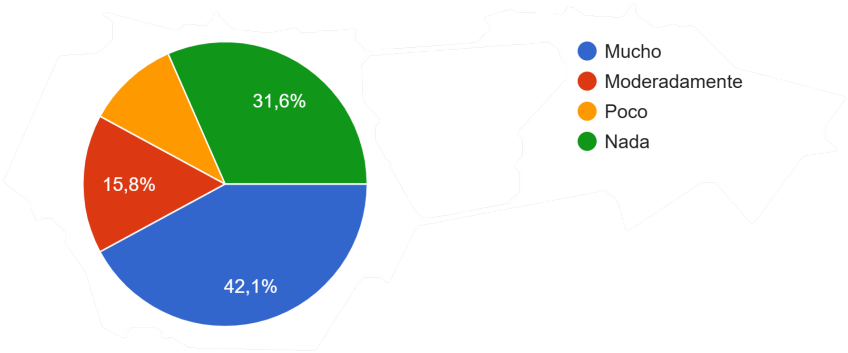
Questions 21 and 22 investigated self-censorship of violent communication, understood as any form of expression that incites physical violence against individuals or groups, that which promotes hate speech or racial discrimination, as well as explicit graphic content or even communications with malicious misinformation. Figures 21 and 22 illustrate the data obtained, in which an average of 68.5% and 57.9% of respondents recognise that they self-censor content published on Instagram and TikTok, respectively, that could be classified by the social network as a violent message.

Figure 21. To what extent do you self-censor violent communication on Instagram?



Source: Own elaboration, 2025.

Figure 22. To what extent do you self-censor violent communication on TikTok?

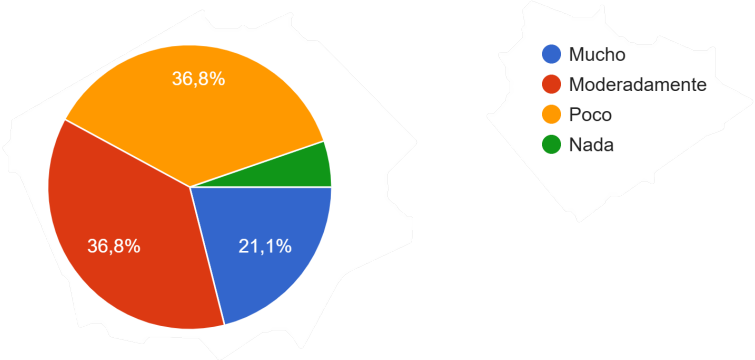


Source: Own elaboration, 2025.

The final questions of the questionnaire, numbers 23 and 24, enquired about students' practices in attempting to circumvent censorship in general terms. The data presented in figures 23 and 24 below reveal that 5.3% of respondents indicated that they do not attempt to circumvent censorship on Instagram, while 26.3% of respondents reported attempting to do so on the social network TikTok. The proportion of respondents who expressed a strong opinion, categorised as 'A lot' or 'Moderately', was 57.9% on the Instagram network, while on the TikTok network, this proportion was 42.1%, representing a 15.8 percentage point difference. This finding suggests that respondents' increased engagement in circumventing censorship on Instagram relative to TikTok may be attributable to a perceived higher level of censorship on the Instagram network.

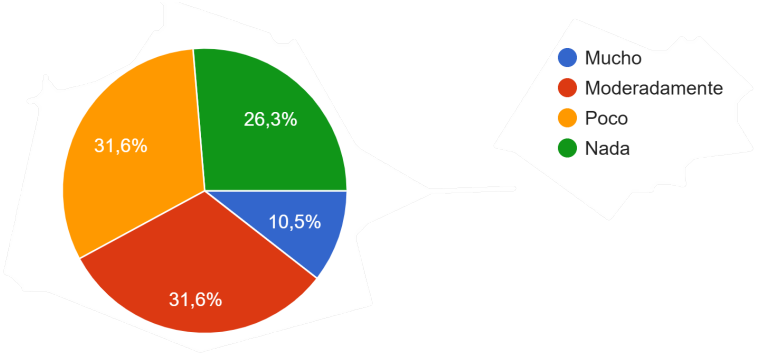
This question is considered to be pivotal, as it has the potential to elucidate significant aspects of the subject under investigation. In the event that users of a given platform engage in practices intended to circumvent the censorship of their content, it can be deduced that they are experiencing heightened pressure in this regard, and that their behaviour is demonstrably influenced by the censorship policies of the network. In this regard, it can be concluded that Instagram exerts a greater degree of influence on the perception of censorship among young students compared to TikTok.

Figure 23. To what extent do you attempt to evade the censorship activities of the Instagram platform?



Source: Own elaboration, 2025.

Figure 24. To what extent do you attempt to evade the censorship activities of the TikTok platform?



Source: Own elaboration, 2025.

4. Discussion and Conclusions

The present study was conducted with the explicit objective of ascertaining the degree of perception of censorship of one's own content in publications and self-censorship on the social networks Instagram and TikTok among university students in the Community of Madrid in December 2024. Consequently, a secondary objective was established with the aim of analysing the behavioural patterns exhibited by these users regarding the publications made on the aforementioned social networks.

Some key aspects of the content publication policies of both social networks have been found to share some commonalities, including aspects such as copyright, disinformation, and privacy treatment. However, it is important to note that the present investigation has been conducted within the regulatory framework for the publication and management of content classified as inappropriate. It is crucial to emphasise the role of algorithms in detecting and moderating content, as they play a pivotal part in the censorship process. It is acknowledged that companies do not disclose the operational principles of their content moderation algorithms; however, certain aspects can be deduced from their usage policies, in conjunction with the outcomes observed from their users' experience in disseminating content, as outlined in Section 3 of this research. A notable finding is that 47.4% of survey respondents reported experiencing censorship on Instagram, while 36.9% reported experiencing it on TikTok. Of these respondents, 5.5% indicated that they had encountered a high degree of content moderation and censorship on both platforms. This content, in the publication stage, is refined by moderation algorithms, defined as a set of defined, systematic, ordered and finite instructions determined in the following axes.

It can be concluded that the algorithms of the two social networks under scrutiny are programmed to identify keywords, phrases and hashtags associated with content that is deemed inappropriate, violent, discriminatory or in contravention of community rules. Furthermore, techniques based on artificial intelligence are employed for the recognition of images and videos with the aim of detecting explicit or violent content or content that promotes illegal activities, to be censored. Furthermore, it has been determined that user complaints constitute a significant signal for the algorithms. Specifically, it has been observed that posts receiving numerous complaints are more likely to be reviewed by a human moderator and are also more likely to be deleted. The automatic collection of user activity by platforms engenders behavioural patterns that can be utilised by the algorithms to identify suspicious behaviour, such as accounts that repeatedly post inappropriate content or engage in bullying.

In order not to distract from the object of study and understanding that the reader can draw their own conclusions from the fieldwork carried out in the present investigation, we can draw some key conclusions. For instance, a mere 5.3% of survey respondents consider the content moderation policies of both platforms to be very fair and well explained, which, in conjunction with the 57.9% who consider them to be moderately so, amounts to a total of 63.2%. This suggests that young people find them acceptable. However, it is important to note that 36.9% of respondents expressed the contrary opinion, perceiving the content moderation policies to be unfair and poorly explained. This suggests a potential area for improvement for both platforms, namely enhancing the transparency of information regarding content restriction.

Furthermore, it can be concluded that the perception of censorship is perceived in a very similar and balanced way on both platforms, with respondents' answers falling in the middle range of "Moderately" on the one hand, and the sum of "Little" and "Nothing" on the other. The perception of censorship did not attain a high rating on either of the two platforms, thus leading to the conclusion that the experience of censorship is average and acceptable.

The temporal dimension of censorship, defined by the study's examination of experiences before, during, and after publication, also falls within the middle range on both social networks and across all proposed times, suggesting that the temporal variable does not exert a statistically significant influence on the experience of censorship on the two platforms under study.

Furthermore, the data obtained indicates that users have not significantly altered their behaviour when generating content based on prior experiences of censorship on either of the two networks. The findings do not allow for a definitive conclusion regarding the potential implications of prior censorship experiences on user behaviour. It remains uncertain whether users exhibit a disregard for previously encountered censorship, leading to a perceived indifference towards such restrictions, or whether users, despite the presence of censorship, opt to disseminate content that may be considered objectionable, driven by a preference for content that, although subject to subsequent censorship, is uploaded regardless.

The variables of gender, political ideology, xenophobia, body and violent communication have obtained very similar measurements both for Instagram and for TikTok. It is notable that xenophobia is the variable for which users experience the least censorship, which can be understood to mean that it is non-existent (otherwise it would be experienced), although a percentage of 10 is recorded.5% for Instagram and 5.3% for TikTok. However, 5% for Instagram and 5.3% for TikTok reported experiencing significant censorship in this area, suggesting that a relatively small proportion of the audience either disseminates unacceptable content or belongs to the population under study who engage in this type of communication. The remaining variables offer a range of conclusions, which readers can examine in detail for each of the items studied.

In conclusion, it is evident that Instagram users demonstrate a higher propensity to circumvent censorship of their content compared to TikTok users, suggesting a heightened level of censorship pressure on the former platform. Additionally, it can be concluded that both social networks are perceived, in general, in very similar terms in all other aspects studied. The results indicate a unanimous perception of censorship and a response of self-censorship among young people, but manifesting itself, in both cases, with moderate rates, both for Instagram and for TikTok.

References

Balkin, J. M. (2004). Virtual liberty: Freedom to design and freedom to play in virtual worlds. *Virginia law review*, 2043-2098. https://doi.org/10.2307/1515641

Bard, G. & Magallanes, M. L. (2021). Instagram: La búsqueda de la felicidad desde la autopromoción de la imagen. Culturales, 9(1), 519. https://doi.org/10.22234/RECU.20210901.E519

Bilewicz, M. & Soral, W., (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. Wiley Online Library M Bilewicz, W Soral Political Psychology, 2020. Wiley Online Library, 41(S1), 2020. https://doi.org/10.1111/pops.12670

Conde, M. A. (2024). Explorando las tendencias y tácticas de control en internet: un análisis global de los bloqueos y censura en redes sociales. *Revista de Comunicación de la SEECI*, *57*, 1-19. https://doi.org/10.15198/seeci.2024.57.e870

Evolvi, G. (2019). Islamexit: inter-group antagonism on Twitter. Information, Communication & Society, 22(3), 386-401. https://doi.org/10.1080/1369118X.2017.1388427

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moder ation, and the Hidden Decisions that Shape Social Media*. Yale University Press.

Gomes-Franco-Silva, F. & Sendín-Gutiérrez, J. C. (2014). Internet como refugio y escudo social: Usos problemáticos de la Red por jóvenes españoles. *Comunicar: Revista Científica de Comunicación y Educación*, *22*(43), 45-53. https://doi.org/10.3916/C43-2014-04

Gorwa, R., Binns, R. & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, *7*(1). https://doi.org/10.1177/2053951719897945

Heath, A. (2022, junio 15). Facebook plans 'discovery engine' feed change to compete with TikTok | The Verge. https://www.theverge.com/2022/6/15/23168887/facebook-discovery-engine-redesign-tiktok

Instagram (2025). Normas comunitarias. Servicio de ayuda de Instagram. https://es-es.facebook.com/help/instagram/477434105621119/?helpref=hc_fnav

Jansen, S., & Martin, B. (2015). The Streisand effect and censorship backfire. *International Journal of Communication*, 9, 656-671. https://1932--8036/20150005

Llansó, E., Hoboken, J., Leersen, P. y Harambam, J. (2020). *Artificial Intelligence, Content Moderation and Freedom of Expression*. Transatlantic Working Group.

MacKinnon, R. (2012). Consent of the networked: The worldwide struggle for Internet freedom. *Politique étrangère*, *50*(2), 432-463. https://dl.acm.org/doi/abs/10.5555/2530486

Martínez-Sánchez, M. E., Vicario, L. M., y Nicolas-Sans, R. (2022). Censura y Redes Sociales: El caso de Zahara. *VISUAL REVIEW. International Visual Culture Review/Revista Internacional de Cultura Visual*, *10*(1), 1-10. https://doi.org/10.37467/revvisual.v9.3554

Martínez-Valerio, L., y Mayagoitia-Soria, A. (2021). Influencers y mensajes de odio: jóvenes y consumo de contenidos autocensurados. *Revista Prisma Social*. https://hdl.handle.net/10115/30379

Pérez, G. E. M. y El Mecky, N. (2024). Un enfoque de derecho contractual para la censura privada del arte en plataformas de redes sociales. *IDP: revista de Internet, derecho y política*, (41), 3. https://dialnet.unirioja.es/servlet/articulo?codigo=9700831

Pérez, R. V., Catalina-García, B. y de Ayala, M. C. L. (2019). Participación y compromiso de los jóvenes en el entorno digital. Usos de las redes sociales y percepción de sus consecuencias. *Revista Latina de comunicación social*, (74), 554-572. https://doi.org/10.4185/RLCS-2019-1345

Pino, E. B., Quero, H. J., & Díaz, G. H. (2022). Esplendores y miserias de las redes sociales. *Comunicación: estudios venezolanos de comunicación*, (199), 127-142. https://dialnet.unirioja.es/servlet/articulo?codigo=8856251

Piñeiro-Otero, T. y Martínez-Rolan, X. (2024). Interacciones digitales y desigualdad de género: Un estudio sobre el uso de Instagram entre alumnado universitario. Gender on Digital. *Journal of Digital Feminism*, 2, 33-56. https://doi.org/10.35869/GOD.V2.5892

Ramírez, J. F., Corchado, D., Morejón, M., Ramírez, J. F., Corchado, D. y Morejón, M. (2021). Algoritmo para la medición y análisis de la autoridad e influencia de los usuarios en las redes sociales y

profesionales. *PAAKAT: revista de tecnología y sociedad*, 11(21), 1-27. https://doi.org/10.32870/PK.A11N21.598

Rosales, M. J. C., Ramírez, G. J. G., Velásquez, M. N. Z., Aguiriano, R. S. S. y Maradiaga, L. A. S. (2024). Libertad de expresión en las redes sociales: análisis de la censura automatizada por inteligencia artificial. *La Revista de Derecho*, 45, 295-310. https://doi.org/10.5377/lrd.v45i1.19390

Sánchez-Holgado, P. y Benito, M. E. R. (2024). Nudes on the Internet: Social Perception of Body Censorship on Instagram and Twitter (X). *International Visual Culture Review. Revista Internacional de Cultura Visual*, 16(8), 107-120. https://doi.org/10.62161/REVVISUAL.V16.5416

San Juan, J. L. (2021). El control de contenidos en las redes sociales: la nueva forma de censura de la era digital. *Ibersid: revista de sistemas de información y documentación*, *15*(2), 23-28. https://doi.org/10.54886/ibersid.v15i2.4736

Serrano, J. (2016). Internet y emociones: nuevas tendencias en un campo de investigación emergente= Internet and Emotions: New Trends in an Emerging Field of Research. *Comunicar: Revista Científica Iberoamericana de Comunicación y Educación. Scientific Journal of Media Education: 46, 1, 2016*, 19-26. http://dx.doi.org/10.3916/C46-2016-02.

Stewart, D. R., & Bunton, K. (2016). Practical Transparency: How Journalists Should Approach Digital Shaming and the Streisand Effect. *U. Balt. J. Media L. & Ethics*, *5*, 4.

Sturges, P. (2010). Misterio y transparencia: el acceso a la información en los dominios de la religión y la ciencia. *Ibersid: revista de sistemas de información y documentación*, *4*, 21-28. http://www.ibersid.eu/ojs/index.php/ibersid/article/viewFile/3863/3643.

Vizcaíno-Laorga, R., Catalina-García, B. y de Ayala-López, M. C. L. (2019). Participation and commitment of young people in the digital environment. Uses of social networks and perception of their consequences. *Revista Latina de Comunicación Social*, (74), 554-572. http://www.revistalatinacs.org/074paper/1345/28en.html