ISSN online 1688-4221

Ciencias Psicológicas July-December 2025; 19(2), e-4543
DOI: 10.22235/cp.v19i2.4543

Trolling on Social Media. Psychological Well-being and Normal, Pathological, and Positive Personality

Trolling en redes sociales. Bienestar psicológico y personalidad normal, patológica y positiva

Trolling nas redes sociais. Bem-estar psicológico e personalidade normal, patológica e positiva

Alejandro Castro Solano[1]
María Laura Lupano Perugini[1]

[1] Consejo Nacional de Investigaciones Científicas y Técnicas; Universidad de Buenos Aires

Received: 03/26/2025
Accepted: 09/12/2025

Correspondence
Alejandro Castro Solano
alejandro.castrosolano@gmail.com

How to cite: Castro Solano, A., & Lupano Perugini, M. L. (2025). Trolling on Social Media: Psychological Well-being and Normal, Pathological, and Positive Personality. *Ciencias Psicológicas*, *19*(2), e-4543. https://doi.org/10.22235/cp.v19i2.4543

Funding: Project UBACyT 20020190100045BA: "Perfil psicológico de los usuarios de internet y redes sociales. Análisis de rasgos de personalidad positivos y negativos desde un enfoque psicoléxico y variables psicológicas mediadoras", Universidad de Buenos Aires.

Data availability:
The dataset supporting the findings of this study is available at https://osf.io/wxp47/?view_only=daed2672d9174f2685905fb3819bfec3

Conflict of interest:
The authors declare that they have no conflicts of interest.

Abstract: Objective. This study examined the relationships among the reception of negative comments on social media; psychological well-being; and normal, pathological, and positive personality traits. Method. A total of 799 social media users residing in Argentina participated (338 men, 461 women; *M* = 39.7 years, *SD* = 13.84). The following instruments were used: the Big Five Inventory, the Mental Health Continuum–Short Form, the Inventory of the Five Personality Continuums–Short Version, and an ad hoc survey on the reception of negative comments on social media. The design was cross-sectional, nonexperimental, and correlational. The results. The frequency of negative comments was less than 10 % across all analyzed social media platforms. The degree of self-perceived distress was significantly associated with active use of social media ($r = .24$, $p < .001$) and with lower levels of psychological well-being ($r = -.30$, $p < .001$). Normal personality traits explained 9 % of the variance in the level of distress; when pathological and positive traits were added, they accounted for an additional 10 % of the total. Conclusions. Distress associated with the reception of negative comments on social media is linked to lower well-being and to a personality profile characterized by neuroticism and negative affect. The inclusion of pathological and positive traits improves the explanation of distress beyond normal traits.

Keywords: trolling; positive personality traits; pathological personality traits; normal personality traits; psychological well-being

Resumen: Objetivo. Este estudio examinó la relación entre la recepción de comentarios negativos en redes sociales, el bienestar psicológico y los rasgos de personalidad normales, patológicos y positivos. Método. Participaron 799 usuarios de redes sociales residentes en Argentina (338 varones, 461 mujeres; *M* = 39.7 años, *DE* = 13.84). Se emplearon los instrumentos: Big Five Inventory, Mental Health Continuum-Short Form, Inventario de los Cinco Continuos de la Personalidad - versión breve y una encuesta *ad hoc* sobre recepción de comentarios negativos en redes sociales. El diseño fue transversal, no experimental y correlacional. Resultados. La frecuencia de comentarios negativos fue inferior al 10 % en todas las redes sociales analizadas. El grado de malestar autopercibido se asoció significativamente con el uso activo de redes sociales ($r = .24$, $p < .001$) y con niveles más bajos de bienestar psicológico ($r = -.30$, $p < .001$). Los rasgos de personalidad normal explicaron el 9 % de la varianza del grado de malestar y, al incorporar los rasgos patológicos y positivos, adicionaron en total 10 %. Conclusiones. El malestar debido a la recepción de comentarios negativos en redes sociales está asociado con un menor bienestar y con un perfil de personalidad vinculado con el neuroticismo y el afecto negativo. La inclusión de rasgos patológicos y positivos mejora la explicación del malestar más allá de los rasgos normales.

Palabras clave: trolling; rasgos positivos; rasgos patológicos; rasgos normales; bienestar psicológico

Resumo: Objetivo. Este estudo examinou a relação entre a recepção de comentários negativos nas redes sociais, o bem-estar psicológico e os traços de personalidade normais, patológicos e positivos. Método. Participaram 799 usuários de redes sociais residentes na Argentina (338 homens, 461 mulheres; *M* = 39,7 anos, *DP* = 13,84). Foram utilizados os instrumentos: Big Five Inventory, Mental Health Continuum-Short Form, Inventário dos Cinco Contínuos da Personalidade- versão breve, e um questionário ad hoc sobre a recepção de comentários negativos em redes sociais. O delineamento foi transversal, não experimental e correlacional. Resultados. A frequência de comentários negativos foi inferior a 10 % em todas as redes sociais analisadas. O grau de mal-estar autopercebido associou-se significativamente ao uso ativo de redes sociais (r = 0,24, p < 0,001) e a níveis mais baixos de bem-estar psicológico (r = -0,30, p < 0,001). Os traços de personalidade normal explicaram 9 % da variância do grau de mal-estar; ao incorporar os traços patológicos e positivos, adicionaram no total 10 %. Conclusões. O mal-estar vinculado à recepção de comentários negativos nas redes sociais está associado a menor bem-estar e a um perfil de personalidade marcado pelo neuroticismo e pelo afeto negativo. A inclusão de traços patológicos e positivos melhora a explicação do mal-estar para além dos traços normais.
Palavras-chave: trolling; traços positivos; traços patológicos; traços normais; bem-estar psicológico

For several years, communication through social media has established itself as a dominant mode. Despite the advantages it may entail, research shows that the use of social media has a darker side (Sands et al., 2020). In this study, we refer to the phenomenon known as trolling, which can be understood not only as disruptive behavior but also as a form of dark leisure, insofar as it involves recreational activity that, although transgressive, is carried out for pleasure and entertainment (Scriven, 2025).

There is no precise definition of the term trolling, and discrepancies can be observed when attempting to define it, sometimes even differing from the general understanding people have of what it means (Ortiz, 2020). Nevertheless, most authors consider it a global term encompassing a spectrum of behaviors and multicausal motivations that are antagonistic, antisocial, or deviant in the context of online behavior (Buckels et al., 2014; Buckels et al., 2018; Hardaker, 2010; Phillips, 2015; Sanfilippo et al., 2018). These behaviors are amplified by the anonymity under which such interventions can occur (Nitschinsk et al., 2023; Suler, 2004). In general, trolls often lack a well-defined intention; their primary goal is merely to annoy and generate interference in communication (Hardaker, 2010). Therefore, if users respond to comments and posts made by trolls, it is highly likely that they will enter into a spiral of unconstructive exchanges (Paakki et al., 2021).

According to some researchers, these kinds of disruptive online behaviors represent only one form of trolling, known as kudos trolling. However, trolling has evolved from provoking others for mutual enjoyment and entertainment to a more abusive, aggressive, and reactive form of behavior that is not intended to be humorous, called flame trolling (Bishop, 2014; Komaç & Çağıltay, 2019; March & Marrington, 2019). There are also other types of users known as silent trolls, who do not engage directly in trolling behavior but instead act as spectators of such online conduct. Similarly, so-called supportive trolls not only observe but also "like" trolling behaviors carried out by others, indirectly encouraging the proliferation of these behaviors (Brubaker et al., 2021; Montez & Kim, 2025; Ubaradka & Khanganba, 2024).

In the political arena, troll farms are well known; these entities function to manipulate voters through posts or comments on social media (Denter & Ginzburg, 2021). Recent research has included the analysis of collective trolling behaviors, driven by anonymity, where the mass repetition of disruptive behaviors (e.g., comments, posts) directed at an individual or group increases the intensity of the harm inflicted (Flores-Saviaga et al., 2018; Sun & Fichman, 2019; Truong & Chen, 2024).

The type of trolling behavior may also vary depending on the social network or community to which the perpetrating user belongs (Fichman & Sanfilippo, 2016). In general, trolling is more frequent in posts or publications that involve politics or current events (Jatmiko, 2024; Seigfried-Spellar & Chowdhury, 2017). Thus, the context of an online discussion often promotes trolling behavior, regardless of users' personal characteristics (Bentley & Cowan, 2021).

There is evidence that public figures (e.g., politicians, artists, etc.) are at greater risk of being threatened and harassed (Akhtar & Morrison, 2019; Hoffmann & Sheridan, 2008a, 2008b; James et al., 2016). However, any user may become the target of such aggressive interventions. According to international figures, 41 % of internet users have personally experienced some form of online harassment, ranging from insults to various forms of abuse (Pew Research Center, 2021). When the profiles of public figures are analyzed, this percentage increases exponentially. For example, in a study

conducted with members of the UK Parliament, 100 % of respondents reported having been attacked by trolls, with male user accounts being the most targeted. In contrast, women experience more aggression of a sexual nature (Akhtar & Morrison, 2019). In a recent survey conducted in Argentina with a sample of 877 women, 33 % reported having suffered some form of aggression through social media, particularly of a sexual nature (e.g., receiving inappropriate content or ridicule and harassment) (Defensoría del Pueblo de la Ciudad Autónoma de Buenos Aires, 2024).

Various international studies have sought to explain the causes and consequences of this phenomenon. Several factors may contribute to a user becoming a troll, one of which may be linked to personality traits.

With respect to normal personality traits, the five-factor model (FFM; Costa & McCrae, 1985) is the framework most commonly employed in research related to internet psychology. With respect to the phenomenon of trolling, international studies have shown negative correlations with the traits of conscientiousness and agreeableness, indicating that trolls are generally unreliable and negligent users (Buckels et al., 2014; March et al., 2023). Similar findings have been reported in Argentina (Lupano Perugini & Castro Solano, 2021, 2023).

To address positive and pathological traits, research conducted in Argentina has employed a locally developed model: the positive personality model (PPM; de la Iglesia & Castro Solano, 2018). This model represents an attempt to integrate the pathological traits proposed in Section III of the DSM-5 (negative affect, detachment, antagonism, disinhibition, and psychoticism) by incorporating positive versions of these traits. The positive traits are serenity, humanity, integrity, moderation, and sightliness. These five positive traits are positioned along the health–illness continuum, constituting an additional pole that extends beyond normality: the positive pole. The aforementioned studies revealed negative correlations between these traits and trolling behavior. Among the pathological traits, disinhibition has emerged as the most influential (Lupano Perugini & Castro Solano, 2021, 2023).

The novelty of the present study lies in analyzing the personality profiles of users who are victims of trolling and how they perceive its impact on their psychological well-being.

Research findings concerning emotional impact are often contradictory (e.g., Frison & Eggermont, 2015; Kraut et al., 2002; Lup et al., 2015; Nie et al., 2015). The well-known internet Paradox, proposed by Kraut et al. (1998), posits that technology designed to foster interpersonal communication ultimately produces the opposite effect when used excessively. However, a subsequent follow-up study revealed that this effect dissipated (Kraut et al., 2002) and that internet use has varying effects depending on users' personality traits. For this reason, it is important to study both types of variables jointly. Although the findings are contradictory, research tends to agree that excessive social media use is associated with increased levels of depressive symptoms and social anxiety (e.g., Blease, 2015; Lupano Perugini & Castro Solano, 2019; Shaw et al., 2015; Stover et al., 2023). A recent review by Zubair et al. (2023) confirmed the link between excessive use and the tendency to experience psychopathological symptoms, adding that the COVID-19 pandemic further intensified the use of digital communication media, fostering the development of such symptoms.

The psychological effects of online aggression have primarily been studied in victims of cyberbullying (children) and cybermobbing (adults). The problem with online aggression is that it spreads more quickly and massively, and aggressive content can remain in cyberspace for a long time, causing even greater harm (Mathew et al., 2019). Research shows that victims may experience depression, anxiety, suicidal ideation, and low self-esteem, among other consequences (Kowalski et al., 2014; Laboy-Vélez et al., 2021; Pacheco, 2022; Tristão et al., 2022). In the case of trolling, perpetrators seek only personal gratification, regardless of the distress inflicted on their victims (Craker & March, 2016; Golf-Papez & Veer, 2017). However, trolling can have severe consequences for victims, who report increased suicidal ideation and self-harming behaviors (Coles & West, 2016).

This study analyzes the relationship between psychological well-being and the reception of negative comments on social media, as well as the role of personality traits in this relationship. In the objectives of this research, the expression negative comments is used because participants were asked whether they had experienced trolling behavior in this manner, as many might not recognize the term or assign it the same meaning it has in the literature on the topic (Ortiz, 2020). Therefore, we adopt a broad definition of trolling that can encompass both ridicule and aggressive comments.

The present study is novel, as previous research has focused mainly on victims of cyberbullying, where the victim is known to the perpetrator and there is a clear intention to cause harm (Sest & March,

2017). In contrast, victims of trolling may be public figures or anonymous users, with no prior knowledge between the victim and perpetrator (Fichman & Sanfilippo, 2016).

In recent years, aspects related to the internet and social media use have also begun to be studied via innovative methodologies such as natural language processing and machine learning. For example, Machova et al. (2022) employed machine learning methods to develop detection models capable of distinguishing between troll users and ordinary users. They also applied sentiment analysis methods to identify the typical emotional tone of troll comments. In another study, Shekhar et al. (2023) employed the self-learning hierarchical long short-term memory (HLSTM) technique, inspired by neural learning models, to classify hateful and trolling comments on social media.

In this research, automated text analysis was used to examine responses to open-ended questions provided by participants who reported being social media users. These responses were collected through a survey administered specifically for this study. This methodology falls within the category of open automated text analysis tools, which employ specialized machine learning algorithms to make sense of large amounts of unstructured data (Iliev et al., 2015). Their use has gained popularity in psychology, particularly for predictive purposes (Yarkoni & Westfall, 2017), such as identifying personality traits through social media posts (Bleidorn & Hopwood, 2019; Park et al., 2015). Traditionally, psychology relies on thematic analysis to extract latent themes from open-text responses provided by users. This method faces the limitation of relying on the subjective judgment of expert raters and is not especially useful when researchers need to analyze large datasets. The computer-based algorithms described here allow for more objective analysis of participant-provided information while maximizing savings in time and cost during data processing (Lamba & Madhusudhan, 2019).

The contribution of this study lies in three main aspects: first, analyzing how users who are victims of comments associated with trolling behaviors react, in contrast to cyberbullying, which has received greater attention in the literature; second, examining the role of less-studied personality traits—both positive and negative—in the perception of well-being impact, compared with more traditional traits; and third, incorporating automated text analysis as an innovative methodological tool.

In light of the above, the following objectives were set: (1) to analyze the frequency of negative comments received by social media users according to the social network used, the source, and type of comment; (2) to investigate the reasons participants believe they receive negative comments, what personal action they take, and what opinion they hold about them; (3) to examine the relationship between the intensity of negative comments received and perceived psychological well-being (personal, emotional, and social); and (4) to assess the predictive capacity of personality variables (normal, pathological, and positive traits) on the degree of perceived impact of negative comments received by users.

Method

This study employed a quantitative, nonexperimental, cross-sectional design with a descriptive–correlational scope.

Participants

The sample was convenient and included 799 social media users (338 men, 42.3 %, and 461 women, 57.7 %), with a mean age of 39.7 years (*SD* = 13.84). A total of 7.1 % (*n* = 57) were foreign nationals residing in Argentina. Most participants were employed (*n* = 629, 78.8 %). With respect to education, 38 % (*n* = 303) reported having completed university or tertiary studies, among whom 12.5 % (*n* = 38) had completed postgraduate studies. Fourteen percent (*n* = 112) reported having completed secondary school. Most participants self-identified as belonging to middle (*n* = 474; 59.4 %) or upper-middle (*n* = 138; 17.3 %) socioeconomic levels. The participants were also asked about the number of hours per day they engaged in online activities requiring connection (1: *I do not use the internet*; 2: *Less than one hour*; 3: *1 hour*; 4: *2 hours*; and so forth up to 24 hours). The average daily connection time was 8.12 hours (*SD* = 4.75).

Materials

Negative Comments on Social Media Survey. A self-developed survey, which is based on the content and format of a previous instrument on negative comments and trolling experiences (Akhtar & Morrison, 2019), was adapted to the local population in terms of terminology and commonly used social

networks. As the questionnaire includes both closed- and open-ended questions oriented toward specific experiences, it is not a standardized scale for measuring latent constructs. For this reason, reliability and validity studies, either in their original or adapted versions, are lacking, as their purpose is exploratory and descriptive, with a focus on the collection of factual self-reported information regarding social media use and the reception of negative comments.

The first section of the survey explored the type of social network used, time spent, and primary activities performed on each network, followed by the negative comments received. Each social network is treated individually, given that most studies focus on the use of a specific network, especially Facebook (e.g., Ellison et al., 2007). For each listed network, participants indicated on a Likert scale: time of use (1: *I do not use it* to 8: *I use it more than 4 hours per day*); whether they received negative comments (1: *Never* to 5: *All the time*); whether they engaged in live streaming, stories, or reels (1: *Never* to 7: *Almost all the time*); the perceived negative impact of comments received for those activities (1: *Not at all* to 5: *Very much*); whether they posted on others' accounts (1: *Never* to 5: *Always*); and the perceived impact of receiving negative comments on those posts (1: *Not at all* to 5: *Very much*). Finally, the participants rated the overall perceived impact of having received negative comments on social media (1: *Not at all* to 5: *Very much*).

In the second section, the participants indicated whether the negative comment came from an identifiable or nonidentifiable account, the content of the comment (e.g., false information, political, physical, racial, or sexual), and the main action they took in response (e.g., I read it and do not respond, I read it and sometimes respond, I read it fully, I always read and respond).

In the third section, three open-ended questions were included: why they believed they received negative comments; what action they usually took after receiving them; and an optional space to provide any further comments on the issue. Finally, sociodemographic data were collected (e.g., sex, age, place of residence, educational level, occupation, and socioeconomic level).

Big Five Inventory (BFI; John et al., 1991; Argentine adaptation by Castro Solano & Casullo, 2001). This instrument consists of 44 items rated on a five-point Likert scale (1: *Strongly disagree* to 5: *Strongly agree*). It assesses five personality traits (extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience). The original authors demonstrated its validity and reliability in adult general populations in the United States. Those studies confirmed the concurrent validity with other recognized personality measures. Studies conducted in Argentina have verified factorial validity for adolescent populations, nonclinical adult populations, and military populations (Castro Solano & Casullo, 2001). In all the cases, a five-factor model explained approximately 50 % of the variance in the scores. For the present sample, internal consistency was as follows: extraversion: $\alpha = .79$, $\omega = .79$; agreeableness: $\alpha = .70$, $\omega = .72$; conscientiousness: $\alpha = .81$, $\omega = .82$; neuroticism: $\alpha = .85$, $\omega = .83$; and openness to experience: $\alpha = .79$, $\omega = .82$.

Inventory of the Five Personality Continuums–Short Form (ICCP-SF; de la Iglesia & Castro Solano, 2023). This instrument was developed for use in the Argentine population and operationalizes the Dual Personality Model, which measures five positive traits (serenity, humanity, integrity, moderation, and sprightliness) and five pathological traits (negative affect, detachment, antagonism, disinhibition, and psychoticism). It contains 55 items rated on a six-point Likert scale (0: *Strongly disagree* to 5: *Strongly agree*). Psychometric studies included exploratory and confirmatory factor analyses; internal consistency analyses (Cronbach's alpha and McDonald's omega); convergent validity studies with measures of positive, pathological, and normal traits; and external validity analyses with relevant indicators (well-being and psychological symptoms). For the present sample, the α values for the trait scales ranged between .76 and .90, and the ω values ranged between .87 and .93.

Mental Health Continuum-Short Form (MHC-SF; Keyes, 2005; Argentine adaptation by Lupano Perugini et al., 2017). This 14-item instrument assesses (a) emotional well-being, defined in terms of positive affect and life satisfaction (hedonic well-being); (b) social well-being (including acceptance, social actualization, contribution, coherence, and integration); and (c) psychological well-being, based on Ryff's theory (1989) (autonomy, mastery, personal growth, positive relations, self-acceptance, and purpose). The items ask how often respondents experienced certain emotions on a Likert scale ranging from 0: *Never* to 5: *Every day*. Validation studies in Argentina confirmed the three-factor structure and its invariance by sex and age. Evidence of convergent and divergent validity has also been obtained (Lupano Perugini et al., 2017). The reliability coefficients for the present sample were as follows:

emotional well-being: α = .84, ω = .85; social well-being: α = .75, ω = .77; and psychological well-being: α = .85, ω = .85.

Procedure

Data were collected by advanced students completing a research practicum at a private university in Buenos Aires, Argentina. Participation was voluntary, and no compensation was provided. Surveys were administered online via the SurveyMonkey platform. On the survey's introductory page, participants were asked to provide informed consent, the anonymity of the data was assured, and it was stated that the information would be used exclusively for research purposes. Data collection was supervised by a faculty researcher.

The study adhered to international ethical guidelines (APA and NC3R) as well as the standards of the National Scientific and Technical Research Council of Argentina (CONICET) for ethical conduct in the social sciences and humanities (Resolution No. 2857/2006).

Data analysis

Descriptive statistics and Spearman's correlations were calculated to examine associations between variables, along with z tests with Holm–Bonferroni correction for comparing proportions. The predictive capacity of normal, pathological, and positive personality traits for perceived impact was evaluated via hierarchical multiple regressions, verifying the basic statistical assumption. Open-ended responses were processed by preprocessing techniques and lexical frequency analysis.

For descriptive statistics, frequencies, correlations, and regression analyses, we used Jamovi software, version 2.2.5 solid, which operates through the R environment.

For automated text analysis, we employed the Quanteda package (version 3.3.0) in R (version 4.3.2) via RStudio (version 2024.04.2+764). Preprocessing included conversion to lowercase; removal of Spanish stop words (Quanteda stopword dictionary); elimination of punctuation, numbers, and special characters; and lexical normalization and lemmatization. Stemming was also applied to group morphological variants of the same word, and unigram tokens were created, excluding higher-order n-grams. No minimum frequency threshold for tokens was applied beyond the preprocessing steps described. Intercoder reliability was not employed, as the analysis was automated.

Results

Analysis of Negative Comments Received, by Platform, Comment Type, and Action Taken

First, the frequency of negative comments received was calculated according to the social network, the origin of the comment (identifiable vs. nonidentifiable accounts), the type of comment (e.g., political, physical appearance, or sexual), and the main action taken in response.

Table 1 presents the proportion of users who reported having received negative comments on each social network, in descending order, along with their corresponding 95 % confidence intervals. Only the percentages and confidence intervals for this group are reported, as they constitute the population of interest for subsequent analyses. Overall, the frequency was low across all platforms, with higher rates on Twitter and Instagram, followed by Facebook and WhatsApp, and levels close to 1 % on YouTube and Twitch.

Table 1

Reception of negative comments by the social media network

Social network	% (*n*)	IC 95 %
X (Twitter)	9.7 (39)	7.45-13.48
Instagram	8.8 (35)	6.42-12.04
Facebook	5.6 (22)	3.71-8.30
WhatsApp	5.5 (22)	3.71-8.30
YouTube	1.2 (5)	0.54-2.96
Twicht	1.2 (5)	0.56-3.60

Table 2 displays pairwise comparisons via the z test for differences in proportions, applying the Holm–Bonferroni correction for multiple comparisons. The results indicate that Twitter presented significantly higher percentages than did Facebook and WhatsApp but did not differ from Instagram. In turn, Instagram registered higher values than YouTube and Twitch did but no differences compared with Facebook or WhatsApp did. Both Facebook and WhatsApp presented higher proportions than YouTube and Twitch did, with no differences between them. YouTube and Twitch did not differ significantly.

Table 2

z tests for differences in the proportions of negative comment reception across social media networks

	X (Twitter)	Instagram	Facebook	Whatsapp	YouTube	Twitch
X	—	0.44 (.658)	2.17 (.030)	2.24 (.025)	5.39 (< .001)	5.39 (< .001)
Instagram		—	1.74 (.082)	1.81 (.071)	5.02 (< .001)	5.02 (< .001)
Facebook			—	0.06 (.952)	3.49 (< .001)	3.49 (< .001)
WhatsApp				—	3.44 (< .001)	3.44 (< .001)
YouTube					—	0.00 (1.000)
Twitch						—

Note. The exact significance values are reported in parentheses.

With respect to the sources of negative comments received, they were distributed similarly between identifiable accounts (n = 368; 53.48 % [95 % CI: 49.68–57.24]) and nonidentifiable accounts (n = 321; 46.52 % [95 % CI: 42.76–50.32]), with no statistically significant differences (z test for difference in proportions: $z = 1.15$, df = 1, $p > .05$).

With respect to content type, more than half of the participants who reported having received negative comments indicated that these comments were related to false information and/or political content. Notably, content categories are not mutually exclusive, meaning that the same individual may have received more than one type of comment. Similarly, a substantial proportion of participants reported not having received negative comments.

For actions taken in response to such comments, among the 319 participants who answered this question, 53.3 % reported that they usually read them without responding, whereas 33.2 % stated that they read them and sometimes respond. The remaining participants reported other actions with lower frequency.

Table 3 presents comparisons by sex. Compared with men, women were significantly more likely to receive comments from nonidentifiable accounts. In addition, women tended to receive negative comments related to false information, sexual aspects, or physical appearance more frequently, whereas men were more likely to receive negative comments related to racial aspects. With respect to actions taken in response to negative comments, women tended to read them without responding at higher rates than men did. No significant gender differences were found in the other types of responses.

Table 3

Source, Type of Comment, and Action Taken, by Gender

	Total % (*n*)	**Male % (*n*)**	**Female % (*n*)**	***z***	***p***	***p (Holm-Bonferroni)***
Source of comment						
Nonidentifiable accounts	46.52(195)	38.5 (75)	61.5(120)	3.21	.001	.005
Type of comment						
False information	33.98(141)	41 (59)	58 (82)	2.02	.043	.043
Political	27.23(113)	52 (59)	48 (54)	.33	.740	.740
Physical appearance	16.14 (67)	35 (24)	65 (43)	2.46	.014	.042
Sexual	14.94 (62)	29 (18)	71 (44)	3.30	.001	.005
Racial	7.71 (32)	72 (23)	28 (9)	2.49	.013	.026
Action Taken						
I read it and do not respond	53.3 (170)	41.7 (71)	58.3 (99)	2.14	.032	.064
I read it and sometimes respond	33.2 (106)	40.5 (43)	59.5 (63)	1.95	.051	.077
I read it fully	11 (35)	42.8(15)	57.2 (20)	.85	.395	.395
I read it and always respond	2.5 (8)	50 (4)	50 (4)	-		

Automated test analysis of negative comments

The participants were asked three open-ended questions aimed at exploring, first, why they believed they received negative comments; second, what actions they took after receiving such comments; and third, they were given the opportunity to provide an open comment on the topic. The data are presented in Table 4.

A content analysis was conducted on the participants' responses. First, sentences are tokenized into words or units of meaning. Preprocessing included lexical normalization, removal of stop words, and semantic grouping of similar terms. If two words had a similar meaning (e.g., *ignore*, *not respond*), they were grouped into the same word category. Word frequencies were then calculated for the most frequent tokens. The same procedure was applied to the three open-ended questions.

Notably, a single participant could contribute multiple tokens within the same response. Therefore, the percentages presented in Table 4 correspond to the relative frequency of each token or lexical category over the total number of tokens identified rather than the number of participants. Categories were constructed by grouping tokens with closely related meanings.

Regarding why people believed they received such comments, the analysis of frequent words suggests that this was due to having different opinions or thoughts about people, society, or specific topics (e.g., due to discrimination, an attempt to bother others, differing opinions, or because what I write bothers them).

Second, in response to the question about the actions most frequently taken upon receiving negative comments on social media, 70 % of the actions consisted of blocking and reporting (e.g., block, report, delete). Twenty percent reported responding and/or contacting the commenter, and the remaining 10 % took no action.

Finally, in the open comments, although the words were less interpretable than in the previous questions, they indicated that such comments impact, affect, or harm people (e.g., they make me angry, they frustrate me and make me feel powerless, I try not to get into a ridiculous fight, they affect me when they are aggressive or violent).

Table 4

Content analysis of open-ended responses on negative comments received

	n	**%**
Why do they receive?		
People	77	40.74
Comments/opinion	47	24.87
Receive	28	14.81
Different	15	7.94
Thought	8	4.23
Society	7	3.70
Topic	7	3.70
What action do you take?		
Block	123	49.2
Report	48	19.2
Respond	26	10.4
Contact	26	10.4
No action	27	10.8
Open Comment		
People/person	66	29.33
Comment	48	21.33
Negative	28	12.44
Receive	27	12.00
Impact	26	11.56
Bad	11	4.89
Someone	11	4.89
Affect	8	3.56

Relationships between the intensity of negative comments, internet use variables, and psychological well-being

Correlations were calculated to analyze the relationships between the degree of perceived impact of negative comments, the time and type of social media use, and psychological well-being.

As shown in Table 5, the perceived impact of negative comments received was associated with the active use of social networks rather than more general use, although the magnitude of these associations was modest. The more negative comments were received in relation to activities such as live streaming, stories, and reels, the greater the degree of perceived impact. Conversely, the impact of comments tended to decrease as psychological well-being—particularly personal well-being and, to a lesser extent, emotional well-being—increased.

Table 5

Correlations between the perceived impact of negative comments, internet use variables, and psychological well-being

	Impacts of comments (Rho Spearman)	***p***
Time spent on social media	.05	.280
Active use of social media (stories, reels)	.12	.030
Feedback (stories, reels)	.20	.001
Active use of social media (posts)	.14	.008
Feedback (posts)	.11	.050
Emotional well-being	-.14	.010
Personal well-being	-.20	.001
Social well-being	-.09	.074
Total well-being	-.17	.002

Note. Significant correlations with a small effect size are shown in bold.

Relationships between the perceived impact of negative comments and personality traits (normal, pathological, and positive)

First, significant positive associations were found with the personality traits of neuroticism and negative affect. Likewise, significant negative associations were observed with agreeableness, serenity, and sprightliness (Table 6).

Next, a hierarchical multiple regression analysis was conducted to examine whether personality variables (normal, pathological, and positive) contributed to explaining the variance in the perceived impact of negative comments received (Table 7). Normal personality traits were included in the first step, pathological traits were added in the second step, and positive personality traits were incorporated in the third step. No control variables were included in the previous steps.

Before the models were interpreted, the assumptions of the analysis were tested. Graphical inspection of standardized residuals and normality tests indicated an approximately normal distribution without relevant skewness. Homoscedasticity analyses revealed that the variance of the residuals remained constant across the predicted values. In each model, the absence of multicollinearity was verified through the examination of variance inflation factors (VIFs), tolerances, eigenvalues, condition indices, and variance proportions. None of the models had VIF values greater than 5, tolerances less than .10, or condition indices above 30 accompanied by high variance proportions (> .50) in more than two predictors, indicating statistical independence among the variables.

Overall, personality traits explained 19 % of the variance in the perceived impact of the negative comments received. A comparison of the models revealed that including pathological and positive variables improved prediction beyond that provided by normal personality traits alone, which accounted for only 9 % of the variance. Pathology and positive personality traits together contributed an additional 10 % (5 % each).

Table 6

Correlations between the perceived impact of comments and personality variables

	Impact of comments received (Rho Spearman)	***p***
Normal traits		
Extraversion	.067	.213
Agreeableness	**-.104**	.052
Conscientiousness	.071	.186
Neuroticism	**.256**	.001
Openness to experience	.002	.965
Pathological traits		
Negative affect	**.260**	.001
Detachment	.046	.393
Antagonism	.042	.436
Disinhibition	-.018	.743
Psychoticism	.043	.429
Positive traits		
Serenity	**-.197**	.001
Humanity	-.015	.775
Integrity	-.084	.116
Moderation	.043	.428
Sprightliness	**-.211**	.001

Note. Significant correlations with a small effect size are shown in bold.

Table 7

Hierarchical multiple regression to predict the perceived impact of negative comments

Variable	*B*	95 % CI para *B*		*SE B*	β	*p*	R^2	ΔR^2
		LI	LS					
Step 1: Normal traits							.09	
Extraversion	.040	-.274	.354	.13	.110	.846		
Agreeableness	- .040	-.354	.274	.16	−.032	.600		
Conscientiousness	.152	-.162	.466	.15	.057	.355		
Neuroticism	.608	.294	.922	.11	.302	.000		
Openness to experience	-.040	-.354	.274	.16	−.006	915		
Step 2: Normal traits + Pathological							.14	.05***
Extraversion	-.040	-.354	.274	.10	−.038	-.602		
Agreeableness	-.152	-.466	.162	.09	−.067	-1.06		
Conscientiousness	-.112	-.426	.202	.15	−.037	-.505		
Neuroticism	.208	-.106	.522	.11	.172	.171		
Openness to experience	.152	-.162	.466	.17	.018	.322		
Negative affect	.646	.313	.979	.13	.288	.000		
Detachment	-.400	-.733	-.067	.12	−.190	.010		
Antagonism	.112	-.202	.426	.11	.045	.427		
Disinhibition	-.272	-.586	.042	.15	-.120	.087		
Psychoticism	-.208	-.522	.106	.12	-.084	.141		
Step 3: Normal traits + pathological + positive							.19	.05***
Extraversion	-.040	-.354	.274	.16	−.009	.892		
Agreeableness	-.152	-.466	.162	.12	−.062	.356		
Conscientiousness	.040	-.274	.354	.13	.025	.746		
Neuroticism	.208	-.106	.522	.11	.129	.136		
Openness to experience	.040	-.274	.354	.15	.003	.955		
Negative affect	.476	.143	.809	.17	.271	.001		
Detachment	-.646	-.979	-.313	.17	−.275	.000		
Antagonism	.040	-.274	.354	.14	.030	.618		
Disinhibition	-.272	-.586	.042	.14	-.120	.090		
Psychoticism	-.112	-.426	.202	.13	-.033	.572		
Serenity	.040	-.274	.354	.12	.022	.744		
Humanity	.112	-.202	.426	.11	.048	.458		
Integrity	-.288	-.602	.026	.11	-.115	.054		
Moderation	.272	-.042	.586	.14	.116	.070		
Sprightliness	-.646	-.979	-.313	.10	-283	.000		

Note. CI: confidence interval; LL: lower limit; UL: upper limit. The β values represent standardized coefficients.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

This study examined the relationships among the reception of negative comments on social media; psychological well-being; and normal, pathological, and positive personality traits.

First, the frequency of negative comments received was explored according to the social network used, their origin (identifiable vs. nonidentifiable accounts), and type of comment (e.g., political, physical, or sexual) (Objective 1). The results indicated that the frequency of negative comments received was low (below 10 %). This figure is lower than those reported by the Pew Research Center (2021), who reported that approximately 41 % of users experience some type of harassment through digital media. Cultural differences may influence these discrepancies, suggesting the value of comparing samples from culturally distant countries to identify whether certain contexts foster such online behaviors.

An important aspect, as noted by Zubair et al. (2023), is that the COVID-19 pandemic led many individuals who previously did not use social networks to begin doing so, increasing the risk of overexposure and the likelihood of becoming victims of trolling. During that period, social networks and

other digital media became sources of misinformation and malicious campaigns, which spurred the development of systems designed to detect trolling (e.g., TrollHunter, Jachim et al., 2020).

The negative comments received by the participants in this sample occurred primarily on X (Twitter) and Instagram and, to a lesser extent, on Facebook and WhatsApp. More than half of the comments were related to false information and political content. These findings are consistent with international reports (Fichman & Sanfilippo, 2016; Seigfried-Spellar & Chowdhury, 2017). On the one hand, X (Twitter) tends to be the most common platform for posting politically charged attacks (Bishop, 2014; Komaç & Çağıltay, 2019) and, as previously mentioned, is also frequently used to spread false information (Jachim et al., 2020). Akhtar and Morrison (2019) reported that when politicians' accounts were analyzed, 100 % of them reported being victims of trolling at some point on this platform. Among ordinary users, receiving this type of negative comment is generally linked to responses to their own posts.

Another important aspect of trolling is that it is often carried out from fake accounts or bots (Jiang et al., 2016), especially in relation to political issues, such as during elections with the presence of troll farms (Denter & Ginzburg, 2021). For this reason, recent research has focused on developing machine learning–based techniques to detect troll accounts and differentiate between fake and authentic profiles (Machova et al., 2022). In the present study, half of the negative comments came from nonidentifiable accounts, suggesting that these may correspond to fake accounts, particularly when related to political issues.

In the analyzed sample, women were more likely to receive comments from nonidentifiable accounts. Furthermore, with respect to content type, they tended to receive comments referring to false information and sexual or physical aspects. In contrast, men were more likely to receive negative comments related to racial aspects. These findings are consistent with both international evidence (Akhtar & Morrison, 2019) and local data (Defensoría del Pueblo de la Ciudad Autónoma de Buenos Aires, 2024).

The second objective was to investigate the reasons why participants believed they received negative comments, what actions they took, and what opinions they held about them (Objective 2). An analysis of the participants' open-ended responses revealed that people believed that they received offensive comments because they held opinions and thoughts different from those of others regarding society or current issues, such as politics. This perception aligns with international research on common motives associated with trolling and other antisocial online behaviors (Jatmiko, 2024; Seigfried-Spellar & Chowdhury, 2017).

With respect to actions taken in response to negative comments, women were more likely to read them without responding, unlike men, who were more likely to reply. As described in the literature, trolling is characterized as provocative behavior without any objective purpose other than to disrupt communication (Hardaker, 2010), a phenomenon that is facilitated by anonymity (Nitschinsk et al., 2023; Suler, 2004). Therefore, as Paakki et al. (2021) argue, if users respond to offensive comments, they are likely to become engaged in an exchange that leads nowhere constructive. In this sample, only 20 % admitted to responding to negative comments.

The third objective of this study was to examine the relationship between the intensity of negative comments received and perceived psychological well-being. This aspect is novel because previous studies have analyzed the effects of being harassed online when there is prior knowledge between the victim and perpetrator, which is not the case with trolling (Fichman & Sanfilippo, 2016). In the sample analyzed, an inverse relationship was found between well-being levels and the intensity of negative comments received. In addition, analysis of the open-ended responses revealed that the participants tended to believe that receiving such comments harmed them or had a negative impact. The degree of perceived impact was greater among users who actively engaged in social media by posting or uploading stories or reels. Thus, receiving negative comments when actively producing content was associated with lower levels of perceived well-being. Although these results partially contradict previous findings indicating that the negative psychological impact is greater among passive social media users (Verduyn et al., 2015), they align with other studies reporting that excessive use of social networks is linked to lower well-being (e.g., Blease, 2015; Lupano Perugini & Castro Solano, 2019; Shaw et al., 2015; Zubair et al., 2023).

As noted earlier, studies on the effects of the internet and social media use on well-being tend to yield contradictory results (e.g., Frison & Eggermont, 2015; Kraut et al., 2002; Lup et al., 2015; Nie et al.,

2015). This suggests that other variables may be involved. For this reason, the final objective of this study was to analyze the role of personality traits (normal, pathological, and positive) in the perceived impact of negative comments received. This aspect is also novel, since most previous research has analyzed the personality characteristics of users who perpetrate trolling, not those who are victims (e.g., Buckels et al., 2014; March et al., 2023; Lupano Perugini & Castro Solano, 2021, 2023).

On the basis of the data analyzed, the degree of perceived impact of negative comments received was positively associated with normal trait neuroticism and pathological trait negative affect. It was negatively associated with the normal trait agreeableness and the positive traits serenity and sprightliness. These findings suggest that individuals who feel more affected by negative comments are those with a tendency to experience negative emotions (e.g., anxiety, worry) and a low ability to regulate them. They also perceive difficulties in interacting with others and maintaining clear goals and purposes. Finally, the regression analysis highlighted the need to include pathological and positive traits, not only normal traits, since they add explanatory power to the variance of the perceived impact of negative comments received.

Limitations, future research directions, and practical implications of the findings

First, the main limitation of this study lies in the fact that trolling behaviors were analyzed in a general way on the basis of participants' perceptions of comments that they considered negative. No distinction was made between different types of trolling behavior—whether oriented more toward the perpetrator's simple enjoyment or toward a more aggressive connotation. Future studies could focus on differentiating the types of trolling behaviors, their characteristics, and their consequences.

Another limitation concerns the representativeness of the sample used, not only in terms of size but also in terms of the fact that most participants were residents of urban centers in Buenos Aires, with limited participation from other regions of the country. This sample was also highly educated, which may influence the way individuals interpret receiving negative comments on social media.

Regarding the data collection instruments, most were self-reported inventories, which may have led participants to orient their responses toward socially desirable patterns and avoid disclosing that they had been victims of attacks or ridicule online. These aspects may affect the generalizability of the findings; therefore, it would be desirable to replicate the analyses with more diverse and larger samples.

For future research directions, it would be useful to conduct international comparative studies since, to date, research has focused primarily on individual variables or characteristics of virtual environments without exploring whether trolling has the same prevalence and impact across different sociocultural contexts. Another relevant line of inquiry would be to analyze the frequency and impact of trolling by differentiating the type of social network (e.g., X, Instagram, TikTok), type of user (e.g., politician, artist, influencer, ordinary user), or type of content targeted by aggressive comments (e.g., political, artistic, social posts).

In addition, future studies could examine the relationships between personality traits and the aspects analyzed in the first objective. This would allow researchers to study whether users with certain personality characteristics are more likely to perceive themselves as victims of trolling than other profiles are and how they tend to respond and react to such attacks. Relatedly, it is also important to analyze the role of potential mediating or moderating variables in the relationship between personality factors and the well-being effects of both being a victim of trolling and engaging in trolling behaviors. For example, variables such as exposure time (Alavi et al., 2025) or self-esteem levels may play important roles in this relationship (Zhou et al., 2023).

The findings obtained provide insights that may inform practical implications at different levels and for different stakeholders. At the individual level, it seems important to strengthen factors that can serve as protective agents against disruptive online behaviors, such as self-regulation, moderation, and clarity of goals. This highlights the importance of developing informational campaigns, whether in educational or community settings, that inform users about both the benefits and risks of internet and social media use, as well as how to respond if one becomes a victim of trolling or other forms of online antisocial behavior (e.g., not responding, blocking). The aim is to foster what some authors call digital well-being, understood as a state in which subjective well-being is preserved in an environment characterized by abundant digital communication (Vanden Abeele & Nguyen, 2022). The goal is for individuals to be able to use digital media in ways that promote a sense of comfort, safety, satisfaction, and personal fulfillment.

Finally, at the technological and public policy levels, it is important to continue refining tools that detect troll accounts to take appropriate measures on the basis of the harm they may cause to users. In summary, it is necessary to work toward the development of healthy virtual environments that allow people to access content and interact positively with others—fulfilling the original purpose of internet development.

References

Alavi, M., Garg, A., & Wanigatunga, N. (2025). The relationships between Dark Tetrad traits and adolescent cyberbullying and cybertrolling with online time and life satisfaction as moderators. *Discover Psychology, 5*(1), 29. https://doi.org/10.1007/s44202-025-00351-6

Akhtar, S., & Morrison C. M. (2019). The prevalence and impact of online trolling of UK members of parliament. *Computers in Human Behavior, 99*, 322-327. https://doi.org/10.1016/j.chb.2019.05.015

Bentley, L. A., & Cowan, D. G. (2021). The socially dominant troll: Acceptance attitudes towards trolling. *Personality and Individual Differences, 173*. https://doi.org/10.1016/j.paid.2021.110628

Bishop, J. (2014). Representations of 'trolls' in mass media communication: A review of media-texts and moral panics relating to internet trolling. *International Journal of Web Based Communities, 10*(10), 7-24. https://doi.org/10.1504/ijwbc.2014.058384

Blease, C. R. (2015). Too many 'Friends,' Too few 'Likes'? Evolutionary psychology and 'Facebook Depression'. *Review of General Psychology, 19*(1), 1-13. https://doi.org/10.1037/gpr0000030

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review, 23*(2), 190-203. https://doi.org/10.1177/1088868318772990

Buckels, E. E., Trapnell, P. D., Andjelovic, T., & Paulhus, D. L. (2018). Internet trolling and everyday sadism: Parallel effects on pain perception and moral judgment. *Journal of Personality*, 1-13. https://doi.org/10.1111/jopy.12393

Buckels, E. E., Trapnell, P., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences, 67*, 97-102. https://doi.org/10.1016/j.paid.2014.01.016

Brubaker, P. J., Montez, D., & Church, S. H. (2021). The power of schadenfreude: Predicting behaviors and perceptions of trolling among Reddit users. *Social Media + Society, 7*(2), 1-13. https://doi.org/10.1177%2F20563051211021382

Castro Solano, A., & Casullo, M. M. (2001). Rasgos de personalidad, bienestar psicológico y rendimiento académico en adolescentes argentinos. *Interdisciplinaria, 18*, 65-85.

Coles, B. A., & West, M. (2016). Trolling the trolls: Online forum users constructions of the nature and properties of trolling. *Computers in Human Behavior, 60,* 233-244. https://doi.org/10.1016/j.chb.2016.02.070

Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual.* Psychological Assessment Resources. https://doi.org/10.1037/t07564-000

Craker, N., & March, E. (2016). The dark side of Facebook®: The Dark Tetrad, negative social potency, and trolling behaviours. *Personality and Individual Differences, 102,* 79-84. https://doi.org/10.1016/j.paid.2016.06.043

Defensoría del Pueblo de la Ciudad Autónoma de Buenos Aires. (2024). *Estudio exploratorio sobre la violencia digital con perspectiva de género.* Instituto de Investigaciones Gino Germani, Facultad de Ciencias Sociales, UBA; Iniciativa Spotlight; ONU Mujeres; UNFPA. https://argentina.unfpa.org/sites/default/files/pub-pdf/2024-12/Informe%20-%20versi%C3%B3n%20final.pdf

De la Iglesia, G., & Castro Solano, A. (2018). The Positive Personality Model (PPM): Exploring a new conceptual framework for personality assessment. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.02027

De la Iglesia, G., & Castro Solano, A. (2023). Análisis psicométricos del Inventario de Continuos de la Personalidad, Forma Corta (ICCP-SF). *Interdisciplinaria, 40*(1), 99-114. https://doi.org/10.16888/interd.2023.40.1.6

Denter, P., & Ginzburg, B. (2021). Troll Farms and Voter Disinformation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3919032

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of computer-mediated communication, 12*(4), 1143-1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Fichman, P., & Sanfilippo, M. R. (2016). *Online trolling and its perpetrators: Under the cyberbridge*. Rowman & Littlefield.

Flores-Saviaga, C., Keegan, B., & Savage, S. (2018). Mobilizing the Trump train: Understanding collective action in a political trolling community. En *Proceedings of the International AAAI Conference on Web and Social Media*. https://doi.org/10.1609/icwsm.v12i1.15024

Frison, E., & Eggermont, S. (2015). Exploring the relationships between different types of Facebook use, perceived online social support, and adolescents' depressed mood. *Social Science Computer Review, 34*, 153-171. https://doi.org/10.1177/0894439314567449

Golf-Papez, M., & Veer, E. (2017). Don't feed the trolling: rethinking how online trolling is being defined and combated. *Journal of Marketing Management, 33*(15–16), 1336-1354. https://doi.org/10.1080/0267257X.2017.1383298

Hardaker, C. (2010). Trolling in asynchronous computer mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, *6*(2), 215-242. https://doi.org/10.1515/JPLR.2010.011

Hoffmann, J., & Sheridan, L. (2008a). Stalking, threatening, and attacking corporate figures. En M. Reid, L. Sheridan, & J. Hoffmann (Eds.), *Stalking, Threatening, and Attacking Public Figures: A Psychological and Behavioral Analysis* (pp. 123-142). Oxford Academic.

Hoffmann, J., & Sheridan, L. (2008b). Celebrities as victims of stalking. En M. Reid, L. Sheridan, & J. Hoffmann (Eds.), *Stalking, Threatening, and Attacking Public Figures: A Psychological and Behavioral Analysis* (pp. 195-213). Oxford Academic.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and Cognition*, *7*(2), 265-290. https://doi.org/10.1017/langcog.2014.30

Jachim, P., Sharevski, F., & Treebridge, P. (2020). TrollHunter [Evader]: Automated Detection [Evasion] of Twitter Trolls During the COVID-19 Pandemic. *New Security Paradigms Workshop*, *2020*, 59-75. https://doi.org/10.1145/3442167.3442169

James, D. V., Farnham, F. R., Sukhwal, S., Jones, K., Carlisle, J., & Henley, S. (2016). Aggressive/intrusive behaviours, harassment and stalking of members of the United Kingdom parliament: A prevalence study and cross-national comparison. *Journal of Forensic Psychiatry and Psychology, 27*(2), 177-197. https://doi.org/10.1080/14789949.2015.1124908

Jatmiko, M. I. (2024). Book review: Trolling Ourselves to Death: Democracy in the Age of Social Media, by Jason Hannan. *Television & New Media, 25*(7), 753-756. https://doi.org/10.1177/15274764241261069

Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious Behavior Detection: Current Trends and Future Directions. *IEEE Intelligent Systems, 31*(1), 31-39. https://doi.org/10.1109/mis.2016.5

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory–Versions 4a and 54.* University of California, Berkeley, Institute of Personality and Social Research. https://doi.org/10.1037/t07550-000

Keyes, C. L. M. (2005). The subjective well-being of America's youth: toward a comprehensive assessment. *Adolescent & Family Health, 4*(1), 3-11.

Komaç, G., & Çağıltay, K. (2019). An overview of trolling behavior in online spaces and gaming context. En *2019 1st International Informatics and Software Engineering Conference (UBMYK)*. https://10.1109/UBMYK48245.2019.8965625

Kowalski, R. M., Toth, A., & Morgan, M. (2017). Bullying and cyberbullying in adulthood and the workplace. *The Journal of Social Psychology, 158*(1), 64-81. https://doi.org/10.1080/00224545.2017.1302402

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). Internet paradox revisited. *Journal of Social Issues, 58*, 49-74. https://doi.org/10.1111/1540-4560.00248

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukopadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist, 53*, 1017-1031. https://doi.org/10.1037//0003-066x.53.9.1017

Laboy-Vélez, L., Ríos-Steiner, A. I., & Flores- Suárez, W. (2021). La violencia digital como amenaza a un ambiente laboral seguro. *Fórum Empresarial, 26*(1), 99-112. https://doi.org/10.33801/fe.v26i1.19494

Lamba, M., & Madhusudhan, M. (2019). Mapping of topics in DESIDOC Journal of Library and Information Technology, India: a study. *Scientometrics, 120*(2), 477-505. https://doi.org/10.1007/s11192-019-03137-5

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram #Instasad?: Exploring associations among Instagram use, depressive symptoms, negative social comparison, and strangers followed. *Cyberpsychology, Behavior, and Social Networking, 18,* 247-252. https://doi.org/10.1089/cyber.2014.0560

Lupano Perugini, M. L., & Castro Solano, A. (2019). Características psicológicas diferenciales entre usuarios de redes sociales de alta exposición vs. no usuarios. *Acta Psiquiátrica y Psicológica de América Latina, 65*(1), 5-16.

Lupano Perugini, M. L., & Castro Solano, A. (2021). Rasgos de personalidad, bienestar y malestar psicológico en usuarios de redes sociales que presentan conductas disruptivas online. *Interdisciplinaria, 38*(2), 7-23. https://doi.org/10.16888/interd.2021.38.2.1

Lupano Perugini, M. L., de la Iglesia, G., Castro Solano, A., & Keyes, C. L. M. (2017). The Mental Health Continuum-Short Form (MHC-SF) in the Argentinean context: confirmatory factor analysis and measurement invariance*. Europe´s. Journal of Psycholo*gy, *13*, 93-108. https://doi.org/10.5964/ejop.v13i1.1163

Machova, K., Mach, M., & Vasilko, M. (2022). Comparison of machine learning and sentiment analysis in detection of suspicious online reviewers on different type of Data. *Sensors, 22*. https://doi.org/10.3390/s22010155

March, E., & Marrington, J. (2019). A qualitative analysis of internet trolling. *Cyberpsychology, Behavior, and Social Networking, 22*(3), 192-197. https://doi.org/10.1089/cyber.2018.0210

March, E., McDonald, L. & Forsyth, L. (2023). Personality and internet trolling: a validation study of a representative sample. *Current Psychology*, *43*(6), 4815-4818. https://doi.org/10.1007/s12144-023-04586-1

Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of hate speech in online social media. En *Proceedings of the 10th ACM conference on web science* (pp. 173-182).

Montez, D., & Kim, D. H. (2025). How do silent trolls become overt trolls? Fear of punishment and online disinhibition moderate the trolling path. *Social Media + Society, 11*(1), 1-13. https://doi.org/10.1177/20563051251320437

Nie, P., Sousa-Poza, A., & Nimrod, G. (2015). Internet use and subjective well-being in China. *Social Indicators Research, 132*, 489-516. https://doi.org/10.1007/s11205-015-1227- 8

Nitschinsk, L., Tobin, S. J., & Vanman, E. J. (2023). A functionalist approach to online trolling. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1211023

Ortiz, S. M. (2020). Trolling as a collective form of harassment: An inductive study of how online users understand trolling. *Social Media + Society*, 1-9. https://doi.org/10.1177/2056305120928512

Paakki, H., Vepsäläinen, H., & Salovaara, A. (2021). Disruptive online communication: How asymmetric trolling-like response strategies steer conversation off the track. *Computer Supported Cooperative Work, 30*(3), 425-461. https://doi.org/10.1007/s10606-021-09397-1

Pacheco, J. (2022). Variables asociadas al fenómeno del ciberbullying en adolescentes colombianos. *Revista de Psicología, 41*(1), 219-239. https://doi.org/10.18800/psico.202301.009

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology, 108*(6), 934. https://doi.org/10.1037/pspp0000020

Pew Research Center (2021). *Online Harassment*. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/pi_2021-01-13_online-harrasment_0-01-1-png/

Phillips, W. (2015). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. MIT Press.

Sands, S., Campbell, C., Ferraro, C., & Mavrommatis, A. (2020). Seeing light in the dark: Investigating the dark side of social media and user response strategies. *European Management Journal, 38*(1), 45-53. https://doi.org/10.1016/j.emj.2019.10.001

Sanfilippo, M. R., Fichman, P., & Yang, S. (2018). Multidimensionality of online trolling behaviors. *The Information Society, 34*(1), 27-39. https://doi.org/10.1080/01972243.2017.1391911

Scriven, P. (2025). Online trolling as a dark leisure activity. *Annals of Leisure Research*, *28*(2), 283-301. https://doi.org/10.1080/11745398.2024.2358764

Seigfried-Spellar, K. C., & Chowdhury, S. S. (2017). Death and Lulz: Understanding the personality characteristics of RIP trolls. *First Monday, 22*(11). https://doi.org/10.5210/fm.v22i11.7861

Sest, N., & March, E. (2017). Constructing the cyber-troll: Psychopathy, sadism, and empathy. *Personality and Individual Differences, 119*, 69-72. https://doi.org/10.1016/j.paid.2017.06.038

Shaw, A. M., Timpano, K. R., Tran, T. B., & Joormann, J. (2015). Correlates of Facebook usage patterns: the relationship between passive Facebook use, social anxiety symptoms, and brooding. *Computers in Human Behavior*, *48*, 575-580. https://doi.org/10.1016/j.chb.2015.02.003

Shekhar, S., Garg, H., Agrawal, R., Shivani, R., & Sharma, B. (2021). Hatred and trolling detection transliteration framework using hierarchical LSTM in code-mixed social media text. *Complex & Intelligent Systems, 9*, 2813-2826. https://doi.org/10.1007/s40747-021-00487-7

Stover, J. B., Fernández Liporace, M. M., & Castro Solano, A. (2023). Escala de Uso Problemático Generalizado del Internet 2: Adaptación para adultos de Buenos Aires. *Revista de Psicología, 41*(2), 1127-1151. https://doi.org/10.18800/psico.202302.017

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior, 7*(3), 321-326. https://doi.org/10.1089/1094931041291295

Sun, L. H., & Fichman, P. (2019). The collective trolling lifecycle. *Journal of the Association for Information Science and Technology, 71*(7), 770-783. https://doi.org/10.1002/asi.24296

Tristão, L. A., Iossi Silva, M. A., De Oliveira, W. A., Dos Santos, D., & Da Silva, J. L. (2022). Bullying y cyberbullying: intervenciones realizadas en el contexto escolar. *Revista de Psicología, 40*(2), 1047-1073. https://doi.org/10.18800/psico.202202.015

Truong, D.-H., & Chen, J. V. (2024). Understanding the we-intention to participate in collective trolling on social networking sites: The online disinhibition perspective. *PACIS 2024 Proceedings, 17.*

Ubaradka, A., & Khanganba, S. P. (2024). The differential effect of psychopathy on active and bystander trolling behaviors: The role of dark tetrad traits and lower agreeableness. *Scientific Reports*, *14*(1), 9905. https://doi.org/10.1038/s41598-024-60203-6

Vanden Abeele, M. M. P., & Nguyen, M. H. (2022). Digital well-being in an age of mobile connectivity: An introduction to the Special Issue. *Mobile Media & Communication, 10*(2), 174-189. https://doi.org/10.1177/20501579221080899

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., & Kross, E. (2015). Passive Facebook usage undermines affective wellbeing: Experimental and longitudinal evidence. *Journal of Experimental Psychology, 144,* 480-488. https://doi.org/10.1037/xge0000057

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Zhou, Y., Li, F., Wang, Q., & Gao, J. (2023). Sense of power and online trolling among college students: Mediating effects of self-esteem and moral disengagement. *Journal of Psychology in Africa, 33*(4), 378-383. https://doi.org/10.1080/14330237.2023.2219527

Zubair, U., Khan, M. K., & Albashari, M. (2023). Link between excessive social media use and psychiatric disorders. *Annals of Medicine and Surgery, 85*(4), 875-878. https://doi.org/10.1097/MS9.0000000000000112

Authors' contribution (CRediT Taxonomy): 1. Conceptualization; 2. Data curation; 3. Formal Analysis; 4. Funding acquisition; 5. Investigation; 6. Methodology; 7. Project administration; 8. Resources; 9. Software; 10. Supervision; 11. Validation; 12. Visualization; 13. Writing: original draft; 14. Writing: review & editing.

A. C. S. has contributed in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13; M. L. L. P. in 1, 5, 6, 12, 13, 14.

Scientific editor in charge: Dra. Cecilia Cracco.