DISINFORMATION AND POLARIZATION IN THE ALGORITHMIC SOCIETY

BRANCO DI FÁTIMA [ED]

DISINFORMATION AND POLARIZATION IN THE ALGORITHMIC SOCIETY

BRANCO DI FÁTIMA [ED]

Covilhã, Portugal 2024
Cali, Colombia 2024

Contents

ONLINE HATE, DISINFORMATION AND SOCIAL FRAGMENTATION

Branco Di Fátima
/ LabCom – University of Beira Interior

This is the first book of the **Online Hate Speech Trilogy**. It focuses on the relationship between disinformation, political polarisation, and virtual attacks involving toxic language. These phenomena are much more complex than they may initially appear, and they are strongly interrelated with each other in the algorithmic society.

To begin with, there is no universally accepted definition of hate speech. Its characterisation is a point of intellectual dispute that varies across different worldviews (Müller & Schwarz, 2021). However, more consensually, hate speech can be understood as a verbal or non-verbal attack on an individual or a social group. As a form of violent communicative action, its roots lie in the codes and values of a particular culture (Matamoros-Fernández & Farkas, 2021).

Hate speech becomes especially more complex, comprehensive, and harmful on online platforms. It is omnipresent, interactive, and multimedia in nature (Di Fátima, 2023). 24 hours a day, 7 days a week, it spreads rapidly through the vast networks of the Internet, influencing behaviour far beyond its point of origin, across borders. Haters hide behind the anonymity facilitated by digital technologies and find support from like-minded communities online (Amores et al., 2021).

This book brings together chapters written by 16 authors from 11 universities across the globe, examining hate

speech within each of their unique socio-cultural contexts. To achieve this, they utilize both traditional and digital methods, employing quantitative and qualitative data collected from various digital platforms, including websites, instant messaging apps, and social media.

The authors analyse the deep origins of hate speech and its manifestations online. They highlight attacks against the LGBTQ+ community in Europe, journalists in Asia, and black politicians in Latin America. Additionally, these researchers examine the crucial role of fact-checking agencies in confronting toxic language and combating hate speech.

It can also happen that the material basis for hate speech is disinformation, which is exacerbated in polarised societies (Ndahinda & Mugabe, 2022). This issue has become more prominent in the last decade, largely due to its strong correlation with populism, the rise of the new anti-system far-right, and conspiracy theories. Platform algorithms play a significant role in the disinformation process, particularly by creating echo chambers that reinforce individuals' beliefs and can distort reality (Recuero, Soares, & Zago, 2020).

Haters aim not only to destroy their targets but also to influence the public sphere at deeper structural levels. Previously, attempts to influence public opinion relied on *one-to-many* transmission technologies (Cardoso, Lapa, & Di Fátima, 2016). An individual or small group could have access to the technology for disseminating information to a large audience via broadcasting. However, everything has changed in the algorithmic society (Ndahinda & Mugabe, 2022). Now, hateful messages can be generated and distributed on a large scale by any individual with access to new media. Moreover, a significant portion of the global population already enjoys such access (Reuters Institute, 2024).

This book uncovers the intimate connection between disinformation, polarisation, and virtual attacks in the algorithmic society. Volumes 2 and 3 of the **Online Hate Speech Trilogy** delve into the legal challenges of combating toxic language while ensuring freedom of expression. They also

explore hate as a political strategy, the latest techniques for detecting violent narratives, and the creation of counter-narratives to mitigate hate speech. The aim is to provide a multicultural overview of one of the most pressing issues in contemporary society, which is responsible for undermining democratic values.

References

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P. & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter: Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info*, 49(2021), 98-124. https://doi.org/10.7764/cdi.49.27817

Cardoso, G., Lapa, T., & Di Fátima, B. (2016). People are the message? Social mobilization and social media in Brazil. *International Journal of Communication*, 10(2016), 3909-3930.

Di Fátima, B. (2023). *Hate speech on social media: A global approach*. LabCom Books & EdiPUCE.

Matamoros-Fernández, A. & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media*, 22(2), 205-224. https://doi.org/10.1177/1527476420982230

Müller, K. & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4), 2131-2167. https://doi.org/10.1093/jeea/jvaa045

Ndahinda, F. M. & Mugabe, A. S. (2022). Streaming hate: Exploring the harm of anti-banyamulenge and anti-Tutsi hate speech on Congolese social media. *Journal of Genocide Research*, 1-15. https://doi.org/10.1080/14623528.2022.2078578

Recuero, R., Soares, F., & Zago, G. (2020). Polarização, hiperpartidarismo e câmaras de eco: Como circula a desinformação sobre Covid-19 no Twitter. *SciELO Preprints*. https://doi.org/10.1590/scielopreprints.1154

Reuters Institute. (2024). *Digital News Report 2024*. Reuters Institute for the Study of Journalism.

THE VIRALIZATION OF STIGMA ONLINE: HATE SPEECH AGAINST GAY MEN IN CONNECTION WITH THE MONKEYPOX OUTBREAK

Adolfo Carratalá
Universitat de València, Spain
adolfo.carratala@uv.es

The international outbreak of monkeypox that began in May 2022 placed Spain as the global epicenter of the disease. With the memory of the pandemic caused by COVID-19 still so fresh, the fear of the rampant spread of a virus drew enormous attention from the media. The first news published about the outbreak identified one social group as being principally responsible for the increase in the number of infections: gay men. This chapter presents the analysis of comments posted by Twitter users in response to news published on the monkeypox outbreak in Spain during the month of May 2022. The aim is to observe the relationship between the stigmatization of the gay population in the media discourse and the reaction of the social audience in terms of hate speech, while also detecting whether this relationship is stronger in the case of content published by pseudo-media that spread disinformation. The research undertaken has demonstrated the close connection that can be seen between disinformation, journalistic coverage from a conservative perspective, and hate speech against the LGBTIQ+ community on social networks. Disinformation and stigmatization went hand in hand for weeks, providing the perfect breeding ground for the generation of hate speech. As this study has shown, although the stigmatizing nature of the news was particularly pronounced in the content

provided by a pseudo-media outlet specializing in the production of disinformation, homophobia also extended to the coverage of the events by conservative mainstream media organizations.

MAINSTREAMING THE FAR-RIGHT: DEPLATFORMING, CONTENT MODERATION AND FAR-RIGHT PRESENCE ON MAINSTREAM SOCIAL MEDIA SITES

Sarah Anne Dunne
University College Dublin, Ireland
sarah.dunne1@ucd.ie

Eugenia Siapera
University College Dublin, Ireland
eugenia.siapera@ucd.ie

There has been a noted rise in the (re)legitimation of far-right political ideology across much of the West and parts of the Global South in recent years. This has been especially noted on various social media platforms, both alternative and mainstream. Indeed, in spite of ongoing content moderation and deplatforming efforts, far-right ideologues maintain a notable presence on mainstream social networks and continue to share anti-democratic and hate-fuelled messages across them. This paper highlights how such actors are able to maintain a visible mainstream presence that disseminates such messages despite hate speech policies and user agreements. Focusing on META/Facebook as an example, we provide a visual and discursive analysis of a post that disputes the availability of LGBTQ+ material for minors, highlighting how the far-right employs tactical manipulation and side-stepping of META's hate speech policies to ensure their material remains accessible on the sight. Via the avoidance of naming LGBTQ+ groups specifically, and the means by which they verbally direct and focus their argument on children as symbols of innocence, rather than on the LGBTQ+ community specifically, these far-right ideologues maintain a mainstream presence that furthers their perceived relevance and legitimacy.
Keywords: mainstream social media, alt-tech platforms, content moderation, deplatforming, far-right politics, anti-LGBTQ+ narratives

FREEDOM OF EXPRESSION, DISINFORMATION AND HATE SPEECH: THE CASE OF COLOMBIA'S VICE PRESIDENT, FRANCIA MÁRQUEZ

Lina Uribe-Henao
Universidad Icesi, Colombia
lina.uribe@u.icesi.edu.co

Maria Juliana Soto-Narváez
Universidad Icesi, Colombia
mjsoto@icesi.edu.co

Mauricio Guerrero-Caicedo
Universidad Icesi, Colombia
mguerrero@icesi.edu.co

This chapter addresses the relationship between disinformation, hate speech, and freedom of expression in the Colombian context based on the analysis of contents that circulated about Francia Márquez during her 2022 presidential campaign and the first months as vice-president of Colombia, and which were checked by two digital media: ColombiaCheck and La Silla Vacía. By reviewing the complexities of this case, we seek to continue contributing to the fight against disinformation, understanding its strategic, discursive, and narrative nuances, and knowing the contributions that a transdisciplinary look can make to its theoretical and practical approach.
Keywords: freedom of expression, disinformation, polarization, hate speech, Colombia

HATE SPEECH IN ARGENTINE SOCIETY: ORIGINS AND CONTEMPORARY EXPRESSIONS

Sergio G. Grabosky
University of Salta, Argentina
graboskysergio@hum.unsa.edu.ar

This article studies the historical, discursive, legal, psychological and cultural dimensions of Hate Speech in Argentina from the Complex Analysis of Discourse. Even though it is a contemporary communicational phenomenon driven by the context of viral online speech, its origins in history can be rebuilt. Hatred towards the abominable subject "negro", a term that in Argentina means native, northerner, brown, poor, peronista; and who goes through national history. However, it is the racialization of politics which was originated by the hatred towards peronismo that gave birth to the most representative term of hate speech in politics: cabecita negra. This political and racial hatred not only attacks those sectors that have received State assistance through the implementation of Public Policies by peronismo, planeros or unemployed and informal workers, but also women's struggle and members of the LGBTIQ+ community. In 2022, this political hatred with strong historical roots has led to a hate crime, the attempted magnicide against the vice-president of Argentina. In 2023, a new antidemocratic atmosphere is noticeable in the country, with the aggravation of institutional violence of a State that undermines human rights and citizens in some Argentine provinces. Forty years after the return to Democracy, the strength of some far-right sectors that put democracy in danger through the growth of fascist stands that put forward hatred, segregation and increase of anti-rights policies are also noticed. Hate speech, paradoxically, has been part of the political campaign that helped Libertad Avanza candidate, Javier Milei, to win the elections and be the elected President.

Keywords: hate speech, political hatred in Argentina, online writing, racialization of politics

ACTIVISM AS A RESPONSE TO HATE SPEECH ON SOCIAL MEDIA: IDENTITY AND FANDOM ON THE @PABLLOVITTAR PROFILE

Marcelo Freire
Federal University of Ouro Preto, Brazil
marcelofreire@ufop.edu.br

Debora Cristina Lopez
Federal University of Ouro Preto, Brazil
debora.lopez@ufop.edu.br

Daniel Martín-Pena
University of Extremadura, Spain
danielmartin@unex.es

At the Lollapalooza 2022 festival, Brazilian artist Pabllo Vittar made a political statement in support of Luiz Inácio Lula da Silva, at that time a pre-candidate for Brazil's presidential elections. In this article, we analyse 4,950 comments made on the posts of the @pabllovittar Twitter profile, specifically in relation to the Lollapalooza Festival. We developed a qualitative analysis, with data coding techniques that guide a deductive approach. The goal is to understand how Pabllo Vittar's fandom responds to comments from Bolsonarist groups and triggers mobilization and identification with the artist as a strategy against hate speech. The data shows that the positive mobilization and identification of the fandom with Pabllo Vittar were the main strategy used. The cycle of aggressiveness and polarization characteristic of interactions on social networks was contained by the identification of the fandom with Vittar and by a strategy collectively built to combat hate speech.

Keywords: activism, hate speech, Twitter, Pabllo Vittar, Lollapalooza festival

POLITICS, JOURNALISM AND HATE SPEECH IN PAKISTAN

Imran ul Haq
Bahria University, Pakistan
iamhaq@gmail.com

Shabir Hussain
Bahria University, Pakistan
shasain2@gmail.com

Jamal Ud Din
University of Swat, Pakistan
jamaluddin@uswat.edu.pk

This chapter offers an insight into the patterns of digital hate speech in the polarized politics and the society of Pakistan. We review the relevant academic scholarship in the context of prevalent political situation in the county to account for the broader factors leading to this malaise. The analysis suggests that the polarized and highly volatile political environment is contributing to the increasing incidents of hate speech against media workers. The social media teams of various political parties and the avid supporters do not allow critical journalists to work independently. The social media affordances and the lack of clearcut government policy against hate speech in Pakistan make them easy targets. We broadly contextualize these arguments within the framework of journalistic norms and digital affordability to explore how divergent reporting patterns adopted by journalists may clash with public expectations and sentiments, thus warning them the public wrath and hate.
Keywords: social media, hate speech, online harassment, freedom of expression, journalist expression

FACT-CHECKING ON TWITTER IN THE FACE OF FAKE NEWS: COMBATING HATE SPEECH ON SOCIAL NETWORKS

Enrique Javier Díez-Gutiérrez
University of León, Spain
enrique.diez@unileon.es

María Verdeja
University of Oviedo, Spain
verdejamaria@uniovi.es

Eva Palomo-Cermeño
Rey Juan Carlos University, Spain
eva.palomo@urjc.es

Disinformation has become a real problem in the 21st century. Understood as a process where fake news is mixed with truthful information and where different channels cooperate to enhance the credibility of such malicious accounts, it is increasingly raising concern in our societies. Disinformation can affect our perception of reality and is believed to be causing a worrying 'information epidemic'. This 'infodemic' has turned social networks into one of the most widely used channels of spreading hoaxes exponentially increasing the problem of disinformation, the consequences of which have even been warned about by the European Commission. The purpose of this study was to analyse the fact-checking work on Twitter carried out by a professional from the Pandemia Digital account. Specifically, the analysis was carried out around the controversy of macro-farms, which went viral with the hashtag #MacrogranjasNO by the end of 2021. Based on the macro-farms controversy and the statements made by the Spanish Minister of Consumer Affairs, Alberto Garzón, defending extensive livestock farming, a fact-checking experience is analysed in the face of fake news from the Pandemia Digital Twitter account. A qualitative methodology is used, using an in-depth interview with a professional fact-checker. Regarding how fake news works on social networks, it should be noted that three key aspects interact: 1) the attraction of lies, 2) the link with strong emotions, such as

hate, and 3) viralisation with the help of fake accounts or bots. In this disinformation strategy, the media play a very important role as necessary partners in the process. The research carried out showed that the intention of fake news is to influence the perception of reality and the construction of a polarised socio-political climate that can have very serious effects on offline reality. Critical digital literacy is one of the key elements in combating disinformation.

Keywords: disinformation, fact checking, fake news, network noise, critical digital, literacy

THE VIRALIZATION OF STIGMA ONLINE: HATE SPEECH AGAINST GAY MEN IN CONNECTION WITH THE MONKEYPOX OUTBREAK

Adolfo Carratalá

/ Universitat de València, Spain

The international outbreak of monkeypox that began in May 2022 placed Spain as the global epicenter of the disease. With the memory of the pandemic caused by COVID-19 still so fresh, the fear of the rampant spread of a virus — which, although not unknown, had rarely been transmitted outside Africa — drew enormous attention from the media. The first news published about the outbreak identified one social group as being principally responsible for the increase in the number of infections: gay men. In a context of growing hate speech against the LGBTIQ+ community, especially on the internet, the media coverage functioned as an incentive to perpetuate prejudice and stigma against non-normative sexualities. This chapter presents the analysis of comments posted by Twitter users in response to news published on the monkeypox outbreak in Spain during the month of May 2022. The aim is to observe the relationship between the stigmatization of the gay population in the media discourse and the reaction of the social audience in terms of hate speech, while also detecting whether this relationship is stronger in the case of content published by pseudomedia that spread disinformation.

Media discourse and stigmatization of the LGBTIQ+ community

Throughout history, heterosexist logic has excluded non-normative sexualities from public discourse. The best way to preserve and reinforce heteronormativity as the norm to which sexual and affective relationships must conform has long been to exclude any cultural representation of identities and experiences that would call into question the apparent natural complementarity of man and woman as the only possible model (Tin, 2012). In this way, media discourse, both in journalism and fiction, has relegated LGBTIQ+ realities to a state of total silence for centuries (Albertini, 2012). Denying and ignoring affective-sexual diversity has thus facilitated the consolidation of the hegemony of heterosexuality as the only acceptable model for relationships. Nonetheless, the symbolic violence entailed in the censorship of LGBTIQ+ identities has not been the only form of media violence suffered by this community (Martínez, 2016). Since the period of silencing has ended, media coverage of gay, lesbian, bisexual, and trans people has been subjected to prejudice, stereotypes, and stigmatizing strategies, which are clear examples of the inaccurate representation of these social groups (Ventura, 2019). The visibilization of LGBTIQ+ realities has conformed to a heterosexist perspective, presenting non-normative sexualities as otherness, highlighting lurid and sensationalistic aspects that encourage the presentation of these identities as extraordinary experiences — not because they are exceptional, but because they are strange, peculiar, and removed from the dominant model (Carratalá, 2016).

In particular, media discourse toward gay men has contributed to them being linked with perversion, crime, and disease (Fejes & Petrich, 1993). From their portrayal as sinners and criminals, according to religious and law enforcement authorities, the media began to portray them as sick, thereby validating the psychiatric discourse that viewed homosexuality as a medical problem (Kuhar, 2003). Through these strategies, the media has attempted to stigmatize gay identity by subjecting it to a process of symbolic violence through which social differentiations and inequalities are established (Rodríguez Martínez & García Gadivia, 2006). The term stigma

is understood to mean any attribute recognized in another individual that, as well as making him or her different from others, is deeply discrediting toward that person, thereby contributing to him or her being discriminated against (Goffman, 1963). Stigma, as an expression of "devalued stereotypes," finds its root cause in certain institutional practices, such as the activities of the media as key players in the social construction of reality, and the people affected by it being excluded from "social networks, neighborhoods, labor markets, the law, and politics" (Clair, 2018).

The media narrative during the early years of HIV/AIDS, which came to be referred to as "gay cancer, gay pneumonia, pink cancer, gay plague" (Martínez Nicolás, 1994), was a clear example of this process. The association of the virus and the disease with a supposed gay lifestyle, and therefore the establishment of at-risk groups, was based on a cultural and ideological dichotomy between an "us" (heterosexual) and a "them" — the others — (homosexual) (Cruze Bell, 2006), which at the time fuelled homophobic views (Mangeot, 2012) and encouraged a twofold process of stigmatization of the community (Kowalweski, 1988). This stigmatizing discourse permeated so deeply that it continues to surface in recent journalistic coverage of HIV that is clearly homophobic (Carratalá, 2019).

The pandemic caused by COVID-19 has again shown how the twofold stigmatization of the LGBTIQ+ community operates, again in the context of the outbreak of a hitherto unknown virus. Some political and religious figures did not hesitate to use LGBTIQ+ people as scapegoats as they were "singled out, blamed, abused, incarcerated and stigmatized as vectors of disease," as was denounced by the United Nations Independent Expert on Sexual Orientation and Gender Identity, Victor Madrigal-Borloz (United Nations, 2020), who also highlighted the role that social media played in this process; for example, in Malaysia, a message claiming that COVID-19 was a punishment from God for LGBTIQ+ people went viral on the web and contributed to an increase in negative rhetoric against this community. During those months, the circulation of LGTBIQphobic messages grew throughout Europe and Central Asia (ILGA, 2021). Frequently, the media

were responsible for this, as was the case in Seoul that was denounced by Amnesty International: "In May, a COVID-19 cluster outbreak among club visitors in Itaewon, a nightlife district in Seoul, generated media reports that suggested unfounded links between infections and sexual orientation. Some reports included personal information, such as the age, residence, workplace, occupation and commuting patterns of individuals, impacting on their privacy. The discriminatory reports caused stigma against LGBTI people, many of whom subsequently avoided COVID-19 testing for fear of being 'outed.'" (2021: 329).

The symbiotic relationship between disinformation and hate speech

As demonstrated by the stigmatizing role played by the media in the coverage of the HIV/AIDS and COVID-19 crises, the problems of disinformation and hate speech often exist side-by-side at the core of some messages. The growing interest generated in the social sciences toward the study of the phenomenon of fake news and disinformation, as a result of its escalation globally in recent years, has prompted reflection on the relationship that can be observed between such content and hate speech toward vulnerable groups, which are often the favorite target of platforms specialized in the dissemination of pseudo-news and falsehoods (Palau-Sampio & Carratalá, 2022). Wardle & Derakhshan (2017: 20) proposed a categorization of the types of expressions found in the current information disorder in which they identified hate speech as a form of mal-information ("information that is based on reality, used to inflict harm on a person, organization, or country"), which they distinguished from disinformation ("information that is false and deliberately created to harm a person, social group, organization, or country"). Nevertheless, this classification complicates the observation of the interrelationship between hate speech and lies, since fake news, for example, may be linked to the stigmatization and discrimination of vulnerable social groups (Segura, 2021). In fact, more recent studies (Hameleers et al., 2022) have demonstrated that hate speech is often a distinctive element of disinformation (messages whose deliberate purpose is to mislead

the audience), insofar as stereotypes, prejudice, and dehumanization, as well as a lack of empirical evidence and expert knowledge, are some of its inherent attributes.

The concept of hate speech refers to all types of expression that "spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin" (Keen & Georgescu, 2020). The United Nations (2019) also stresses that this speech habitually uses "pejorative or discriminatory language with reference to a person or a group on the basis of who they are." Analysis of this speech cannot be limited to linguistic aspects but must also consider who makes it, who its recipients are, through what channels it circulates, and what interactions occur in relation to it, as well as its effects and interpretations. This is why Communication Studies must be at the forefront of research into this phenomenon (Paz et al., 2020), which represents an enormous challenge for the recognition of diversity.

As Arcila-Calderón et al. (2022) indicate, "hate speech has alarmingly permeated our society and social media have become the most suitable means of propagation." Indeed, the internet, and particularly social networks, have fostered the dissemination of hate speech, thanks to the interaction, instantaneousness, permanence, and anonymity that characterize communication on the web (Martínez Valerio, 2021; Rivera Martín et al., 2022). For several years, the Council of Europe has been warning about how the online environment facilitates the dissemination of these messages: "online hate speech is tolerated more than offline hate speech, and is less subject to control" (Keen & Georgescu, 2020). In Spain, the internet and social networks are playing an increasingly more prominent role in the dissemination of hate speech. Official data recorded by the Ministry of the Interior (2022) is evidence of this: Of all the incidents identified by the institution in 2021, most took place in the digital sphere (37.73% on the internet and 22.29% on social networks). Following the release of this information, the newspaper

El País warned in one of its editorials on October 5, 2022 that "hate speech poisons coexistence and is the precursor to violence, particularly when, all too often, social networks provide a venue for its instantaneous dissemination in closed and often criminal bubbles." In fact, the increase in the spread of hate speech on social networks such as Twitter is correlated with the growth of hate crimes (Amores et al., 2021).

Hate speech against the LGBTIQ+ community in Spain

One of the social groups that is most affected by hate speech is the LGBTIQ+ community. These are messages founded on LGBTIQphobia, understood as a cultural system that regulates sexuality and supports acts of physical, verbal, and symbolic violence that some people decide to exert on those they differentiate and perceive as non-compliant with the norms of sexual practices and gender roles (Martínez, 2016). This type of speech has a serious effect on this community, as "the fear and intimidation resulting from hate crime and hate speech prevent lesbian, gay, bisexual and transgender (LGBT) persons from participating fully in society" (FRA, 2009).

While it is difficult to know the real extent of this type of speech in the European Union because several countries do not have an accurate record of its scale, in Spain, the Ministry of the Interior has been officially measuring it since 2015. According to the latest known data (Ministry of Interior, 2022), 60 episodes related to hate speech against the LGBTIQ+ population were recorded in 2021, which means an increase of 87.50% in the last year. Hate speech based on sexual orientation or gender identity is the second most frequent in Spain, behind messages of a xenophobic nature, although it is experiencing the highest growth. In fact, it is through such speech that most of the discrimination against the LGBTIQ+ community in Spain is perpetrated today (Rebollo et al., 2018). Furthermore, it should be taken into account that official figures are not entirely representative, as information gathered by social organizations, observatories, and support services suggest that the magnitude of violence against these minorities is much greater than that recorded by the authorities (Carratalá & Herrero-Jiménez, 2019).

Once again, the internet and social networks have emerged as powerful forums for its expansion, which is a real challenge for today's societies (Anti-Defamation League, 2021). The concept "online anti-LGBT+ hate speech" refers to any "online communication or expression which advocates, promotes, or incites hatred, discrimination or violence, against any individual or group, because of their sexual orientation, or gender identity." Among its principal manifestations, the following are highlighted: "grossly offensive posts or comments and trolling (posting offensive, upsetting or inflammatory comments online in an attempt to hurt and provoke a response)" (Galop, 2021), which are often the result of coordinated efforts of accounts linked to the far right across a variety of social networks (Boone, 2022). Nevertheless, the latest report prepared by GLAAD (2022) indicates that, rather than protecting their users, most of the tech companies that own these social networks barely provide any information on how they manage and respond to hate speech and attacks targeting LGBTIQ+ people. In evaluating how safe the main social networks are for this minority, none of them manages to achieve more than 50 out of a possible 100 points: Instagram (48%), Twitter (45%), Facebook (46%), YouTube (45%), TikTok (43%). In addition, according to data from the organization, "40% of all LGBTQ adults, and 49% of transgender and non-binary people, do not feel welcomed and safe on social media."

The way in which social network users participate in the production of hate speech against the LGBTIQ+ community has been the object of recent research. Some studies have focused on comments made on news published on the social network Facebook. Research by da Silva and da Silva (2021), for example, revealed that about 27% of the comments on news items in the state of Mato Grosso do Sul, Brazil, were hate speech, many of which expressed "repulsion for the existence or repudiation of LGBT people's attitudes." Carratalá's (2023) work regarding comments made on Facebook to news items in the Spanish press about COVID-19 also found that a significant proportion of the comments from the social audience echoed prejudices and stigmas about gay and bisexual men, expressing demeaning and insulting

opinions. The reactions of Twitter and Instagram users to news items in the Spanish media linked to the LGBTIQ+ Pride 2021 celebrations have also been analyzed by various studies (Martínez Valerio, 2022; Rivera Martín et al., 2022). According to the results of these investigations, the prevalence of hate speech appears to be lower on Instagram than on Twitter, although discriminatory comments of a derogatory and hostile nature toward the LGBTIQ+ population can be detected on both platforms.

Methodology

In May 2022, the emergence of an outbreak of monkeypox virus outside of Africa, where the disease is endemic, led to a further episode of discriminatory discourse and narratives against the LGBTIQ+ community. In fact, as a significant portion of the first Monkeypox cases were identified among gay, bisexual, and other men who have sex with men, the Joint United Nations Program on HIV/AIDS (UNAIDS) expressed concern that some media reporting and commentary was reinforcing homophobic and racist stereotypes (United Nations News, 2022). The objective of this study is to determine the relationship between stigmatizing media coverage and expressions of hate speech made by the social audience on Twitter by analyzing the publications made by some media outlets during the weeks following the appearance of the first cases in Spain. Furthermore, the study aims to ascertain whether the publication of discriminatory comments is higher among the audience of the media outlets that produce disinformation, and whether its scale differs according to the editorial and ideological policy of the organization publishing the news.

Three media outlets were analyzed: one pseudo-media outlet and two mainstream media outlets. The pseudo-media outlet selected was *Mediterráneo Digital* (@MediterraneoDGT), one of the main sources of disinformation in Spain, which promotes itself with the slogan "Politically Incorrect" and whose content includes fake news and falsehoods about LGBTIQ+ people (Palau-Sampio & Carratalá, 2022). The mainstream media outlets chosen were *El País* (@el_pais), the newspaper with the highest circulation

in Spain, which has a progressive political stance; and *ABC* (@abc_es), a long-standing newspaper that is traditionally a point of reference for conservative Spanish audiences. To establish which articles would be analyzed, in terms of their comments, all the pseudo-news about monkeypox that *Mediterráneo Digital* published during the month of May 2022, a total of six articles, were taken as a reference. Following this identification, articles of related information published on the same dates by *El País* and *ABC* were selected. In five of the six cases, it was possible to find in these newspapers news items similar to the pseudo-information published by *Mediterráneo Digital*. The sixth pseudo-news article published by *Mediterráneo Digital* did not have any correlating content in the selected mainstream media. Table 1 shows the headlines under which the selected articles appeared on Twitter.

Table 1: *Twitter* headlines of the tweets analyzed

	Mediterráneo Digital	**ABC**	**El País**
N1 18/05	Health alert! 23 gay men confirmed infected by "monkeypox"	Madrid investigates 23 cases of monkeypox and fears number is rising	Madrid raises number of suspected monkeypox cases to 23
N2 20/05	Gay sauna in Madrid, world epicenter of monkeypox infection	Madrid closes largest gay sauna in the capital, identified as primary source of monkeypox outbreak	Community of Madrid closes sauna in capital where main outbreak of monkeypox detected. Spain is the country with most positive cases, 30 confirmed
N3 23/05	Festival of 80,000 gays in Maspalomas, focus of the spread of monkeypox	Health Ministry investigates a huge festival in Maspalomas as a hotspot for the spread of monkeypox	Massive party in Gran Canaria, held between May 5 and 15 and attended by nearly 80,000 people, is the second hotspot of the monkeypox outbreak in Spain

N4 25/05	WHO confirms that Spain accounts for one-third of monkeypox cases worldwide	WHO rules out that monkeypox virus has mutated	Health Ministry raises cases of monkeypox in Spain to 51 after 15 new positives confirmed
N5 26/05	It's coming: Spain buys millions of doses of monkeypox vaccines	Health Ministry promises vaccines for monkeypox "within days or weeks"	Health Ministry announces the purchase of thousands of vaccines to treat the outbreak of monkeypox. Recognized cases in Spain rise to 59
N6 31/05	Health Ministry confirms having sex with men in "saunas and gas stations" spreads monkeypox	--	--

Source: Prepared by the author.

By means of a methodological approach that combines quantitative content analysis with discourse analysis, an evaluation is made of the stigmatizing nature of the headlines published by the three media outlets on their Twitter profiles (N= 16), the engagement generated through the number of likes, retweets and comments, along with the presence of hate speech among all the comments made by the platform's social audience (N= 622).

Results

Stigmatization in the news coverage on Twitter

Having gathered the 16 articles that form the corpus, an assessment scale was designed to evaluate the level of stigmatization observable in the various sources of (pseudo)information according to the elements they contain and the references to LGBTIQ+ people they make. Accordingly, in order to measure the degree of discrimination observable in the tweets analyzed, four possible levels of stigmatization were considered:

0) No stigmatization: No element of the post alludes to the LGBTIQ+ community.

1) Low stigmatization: The text accompanying the headline and that which can be read in the tweet includes some reference to LGBTIQ+ people. For example, this occurs in Figure 1, corresponding to N1 published by *ABC*, which reads "The vast majority of those affected are young men who had sex with other men."

Figure 1: Tweet from the *ABC* newspaper that reflects a low level of stigmatization

Source: Twitter.

2) Medium stigmatization: Some reference to LGBTIQ+ people is found in the headline or in the image in the tweet. For example, this is the case in N3 published by the newspaper *ABC* in its Twitter profile, which includes an image related to the Gay Pride Maspalomas celebrations held in May 2022 (Figure 2).

Figure 2: Tweet from *ABC* newspaper that reflects a medium level of stigmatization

Source: Twitter.

3) High stigmatization: Both the headline, or the text, and the image accompanying the tweet explicitly refer to the LGBTIQ+ population. This can be seen, for example, in the tweet that *Mediterráneo Digital* issued to publish N2 (Figure 3).

Figure 3: *Mediterráneo Digital* tweet that reflects a high level of stigmatization

Source: Twitter.

The assessment of the 16 selected tweets, based on this stigmatization scale, reveals significant differences between the different accounts under analysis (Table 2). As such, it is noteworthy, first of all, that the tweets published by the pseudo-media outlet *Mediterráneo Digital* are of a notably discriminatory nature. Five of the six posts exhibit some form of stigmatizing feature. Two of them display a medium degree of stigmatization, and three reveal a high level, implying that both the image and the headline (or the accompanying text) overtly identified gay or bisexual men as being responsible for the origin of the monkeypox outbreak in Spain and its subsequent spread. Only the tweet published in relation to N5, related to the purchase of vaccines against the disease, refrained from incorporating discriminatory elements against this social group. An entirely opposite scenario can be observed in the tweets extracted from the profile of the newspaper *El País*, which did

not publish any content explicitly alluding to the LGBTIQ+ population in its coverage of the monkeypox outbreak. Neither the texts nor the images that this media outlet used in its Twitter reports on the matter exhibited discriminatory features. For its part, the conservative newspaper *ABC* also published content with stigmatizing features, of a low or medium level, in three of the five tweets analyzed. They correspond, moreover, to the first reports that the newspaper published to inform about the outbreak of monkeypox cases in Spain. Its last two tweets dropped any reference to gay and bisexual men, both in their texts and their images.

Table 2: Level of stigmatization of the tweets analyzed in each of the news items

| News item | Level of stigmatization | | |
	Mediterráneo Digital	*ABC*	*El País*
N1	2	1	0
N2	3	2	0
N3	3	2	0
N4	2	0	0
N5	0	0	0
N6	3	-	-

Source: Prepared by the author.

Discriminatory speech and social audience engagement

Before analyzing the comments made by the social audience in response to the Twitter posts of the accounts analyzed, it is relevant to quantitatively evaluate the reaction generated by each of the tweets among Twitter users — in other words, what is known as engagement. In this sense, the total number of interactions that each of the news items provoked among

the Twitter audience was computed, adding, consequently, the number of retweets that each publication obtained, the likes it received, and also the number of comments it generated (Table 3).

The pseudo-media outlet *Mediterráneo Digital* achieved a total of 278 reactions, which in most cases took the form of likes (153), retweets (75), and to a lesser extent, comments (50). The tweets that received the most reactions were those related to N2 (65 reactions), N3 (58), and N6 (53). The engagement of these three objects of analysis represent 63% of the total number of reactions achieved by this account in connection with its tweets on monkeypox. It is noteworthy that the three tweets most successful in provoking reactions from the audience correspond to level 3 of stigmatizing speech, the highest level. It can be observed, therefore, that the reaction in terms of retweets, likes, or comments seems to be related to the stigmatizing nature of the text: the higher its level, the more the audience reacts.

Table 3: Level of engagement associated with the tweets analyzed for each of the news items.

	Mediterráneo Digital (75,618 followers)			*ABC* (2,267,008 followers)			*El País* (8,709,862 followers)		
	Retweets	Likes	Comments	Retweets	Likes	Comments	Retweets	Likes	Comments
N1	9	23	11	81	29	13	1906	318	102
N2	22	33	10	1394	1876	239	188	182	57
N3	15	35	8	20	14	3	268	201	80
N4	10	29	12	4	7	2	137	60	12
N5	4	4	0	14	16	28	62	77	36

N6	15	29	9	-	-	-	-	-	-
Total	75	153	50	1513	1942	285	2561	838	287
Media	12.5	25.5	8.3	302.6	388.4	57	512.2	167.6	57.4
Engagement	278			3740			3686		
Standardized Engagement	0.0037			0.0016			0.0004		

Source: Prepared by the author.

A similar relationship can be observed in the case of the *ABC* newspaper profile. This was the media outlet with the highest number of reactions of the three accounts analyzed (3,740 interactions), despite the fact that it has fewer followers on Twitter than the newspaper *El País*. As was the case with *Mediterráneo Digital*, the interaction from Twitter users was focused on liking the tweets of this media outlet (amounting to 1,942 in total), followed by sharing the posts (1,513 retweets) and replying to them (285 comments in total). The tweets that obtained the highest engagement were those related to N2 (3,509 reactions), N1 (123), and N5 (58). As with *Mediterráneo Digital*, two of these tweets demonstrated some level of stigmatization, ranging from low (N1) to medium (N2). It should be noted that the *ABC* tweet related to N2 (Figure 4) is that which obtained the highest number of reactions among the entire sample under analysis, mostly due to the high number of likes and comments it received (only the *El País* tweet related to N1 surpassed it terms of the number of retweets). In this report, the conservative newspaper explicitly linked the outbreak of monkeypox with a gay sauna in the capital of Spain, which it defined as "the focal point" of the virus. In the text accompanying the headline, it was emphasized that the premises that had been closed is "a known venue for sexual encounters between men."

Figure 4: *ABC* newspaper tweet linked to N2, emphasizing the link between monkeypox and a gay sauna in Madrid

The tweets published on Twitter by the account for the newspaper *El País* obtained a high number of reactions (3,686), almost the same as that of the newspaper *ABC*. In this case, in contrast to the other two profiles, most of the engagement was achieved through retweeting the news items (2,561) — this being far higher than the number of likes (838) or comments (287) received. The tweet this newspaper published in relation to N1 was the one which generated the greatest number of reactions in the social audience of Twitter (2,326), primarily due to the large number of users who decided to share this first piece of information regarding the detection of monkeypox virus in a number of Spanish patients (1,906).

If we compare the average number of retweets, likes, and comments that the three accounts received in the group of tweets analyzed, we can see that

the newspaper *El País* managed to achieve more users sharing its tweets (512.2 retweets on average per item published), while *ABC* stands out for being the media outlet with the highest average number of likes (388.4). Both were almost equal in their ability to generate comments from their audiences (57.4 comments on average per tweet in the case of *El País*, and 57 in the case of *ABC*). Evidently, the tweets published by *Mediterráneo Digital* achieved far more modest figures in all the engagement indicators, largely because its number of followers (75,618) is much lower than those of the two other accounts analyzed (*ABC* has 2,267,008 followers, and *El País* has 8,709,862). However, if the interaction generated by the three profiles is standardized by dividing the total number of reactions by the number of followers of each account, we can see that the level of engagement is higher in *Mediterráneo Digital* (0.0037) than in *ABC* (0.0016) and *El País* (0.0004). The reaction of the social audience appears to be directly proportionate to the stigmatizing nature of the accounts analyzed, with the pseudo-media outlet, known for spreading disinformation, managing to generate the highest number of interactions with its posts.

Hate speech in the social audience comments

After evaluating the stigmatizing nature of the publications that constitute the corpus and measuring the engagement associated with them, it is essential, within the framework of the objectives of this study, to further examine the comments made by the social audience of the three accounts to identify the presence of hate speech among the opinions that Twitter users expressed in relation to these news items.

Graph 1 illustrates the result of this analysis in global terms for each of the accounts considered. As can be seen, the profile with the highest number of comments regarded as hate speech toward the LGBTIQ+ community in general, particularly against gay and bisexual men, was that of the newspaper *ABC* (70) — far ahead of those for the newspaper *El País* (48), despite the fact that both had a very similar number of total comments, and those of the pseudo-media outlet *Mediterráneo Digital* (29).

However, if we look at the proportion of hate speech in relation to the total number of comments made for each of the profiles, we can see that the *Mediterráneo Digital* account had the highest relative percentage of hate speech in connection with its tweets. More than one out of every two comments posted by the social audience to the pseudo-news published on this account (58%) corresponds to this type of speech. This percentage is lower in the case of the newspaper *ABC*, accounting for 25% of the total number of comments, and even lower in the profile of the newspaper *El País*, where hate speech is observed in only 17% of the opinions expressed by the users of the social network who decided to comment on its tweets. Therefore, it can be considered that the presence of hate speech is greater in the pseudo-media outlet that specializes in producing disinformation than in the mainstream media. Furthermore, when comparing the latter two newspapers, it can also be concluded that the expression of hate speech is more closely linked to the newspaper that reflects a conservative editorial position (*ABC*) than to the one that is noted for upholding more progressive ideals (*El País*).

Graph 1: Presence of hate speech in the social audience comments of the three profiles
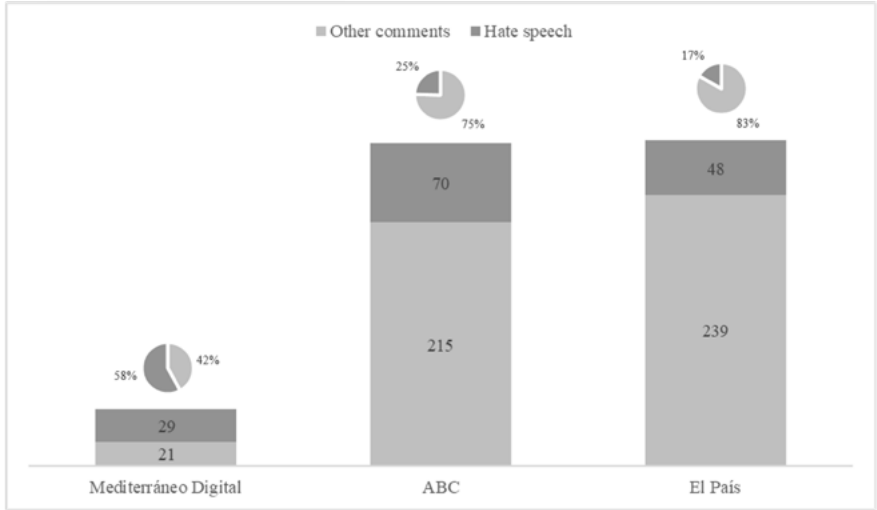
Source: Prepared by the author.

If instead of evaluating the presence of hate speech in the total number of comments received by each of the accounts we consider what proportion they represent of the opinions expressed in relation to each of the news items analyzed (Table 4), some interesting observations can be made. First, it should be noted that all the tweets analyzed received responses that can be considered as hate speech, with the exception of three posts: the N5 tweet from *Mediterráneo Digital* and the tweets linked to N4 published by *ABC* and *El País*. The common characteristic of these tweets is that none of them featured stigmatizing elements toward the LGBTIQ+ community. This means that 81% of the tweets analyzed included hate speech.

Table 4: Presence of hate speech in the comments to the tweets analyzed for each of the news items

	Presence of hate speech in the comments to news items		
	Mediterráneo Digital	*ABC*	*El País*
N1	8 (73%)	2 (15%)	2 (2%)
N2	4 (40%)	64 (27%)	12 (21%)
N3	5 (63%)	2 (67%)	33 (41%)
N4	8 (67%)	0	0
N5	0	2 (7%)	1 (3%)
N6	4 (44%)	-	-
Total	29 (58%)	70 (25%)	48 (17%)
Media	4.8	14	9.6

Source: Prepared by the author.

If we examine the figures associated with the pseudo-media outlet *Mediterráneo Digital*, we can see that the number of comments considered as hate speech is proportionally very high; for example, they amount to 73% in the case of the tweet linked to N1 ("Health Alert! 23 gays confirmed

infected by 'monkeypox'") and 67% in those linked to N4 (Figure 5). In the case of the comments accompanying the tweets published by the *ABC* newspaper, which is the media outlet with the highest total number of comments of the three profiles examined, it can be seen that it is mainly the tweet relating to N2 (Figure 4) that accounts for the highest number of messages considered as hate speech (91% of all those identified in this case). It is also the thread with the greatest number of accompanying reactions from the social audience in which such expressions occur, it being the post in which the newspaper established a direct link between the emergence of the monkeypox outbreak and a sauna for gay and bisexual men in Madrid. The remaining tweets published by this newspaper received a very low number of reactions that could be considered as hate speech, with no more than two comments for each news item.

Figure 5: Tweet from the pseudo-media outlet Mediterráneo Digital associated with N4

Source: Twitter.

In the case of the newspaper *El País*, the highest proportion of hate speech found in the responses among all its tweets was 41% of the total number of comments received. This percentage was observed in the publication linked to N3, on the association between the outbreak of the virus and the Maspalomas Pride celebration (although the newspaper did not explicitly identify this event in its post). In addition to this tweet, the one related to the closure of the sauna in the capital, a place that the newspaper never described with any particular adjectives or linked to any specific social group, also stands out for receiving a significant percentage of comments in which hate speech is detected (21%). The remaining news items published in the profile of *El País* hardly received any reactions that could be considered to display hatred toward the LGBTIQ+ community.

The identification of the news items in which the occurrence of hate speech was highest makes it possible to establish a close relationship between the generation of stigmatizing content toward gay men and the expression of prejudiced, hostile, and discriminatory opinions toward this community. Accordingly, the four tweets which received reactions expressing hate speech in more than 50% of the comments made by the social audience (MD-N1, MD-N3, MD-N4, ABC-N3) correspond to publications which, according to the rating scale used in this study, exhibited a medium or high level of stigmatization. Similarly, it is also evident that the clearly disinformative content (most of these tweets were published by the pseudo-media outlet *Mediterráneo Digital*) encouraged the production of hate speech by Twitter users.

Detailed observation of the comments considered to be examples of hate speech makes it possible to establish, inductively, three different categories into which these remarks can be classified. In first place, some comments can be detected that identify gay men as being responsible for the outbreak and propagation of the virus in Spain, thus reinforcing an association suggested by some of the tweets analyzed. These are expressions of hatred that make a direct link between monkeypox and the sexual orientation of these people, even suggesting that measures should be taken against them

to protect the rest of the population. This finger-pointing can be observed in opinions such as the following: "The fag virus," "Mass gay lockdown please. Let's not play with the health of innocent people," "Thanks to all those degenerate Gays now the whole country is on health alert," "It all started with a faggot fucking a monkey I SCREAM," "They should just start closing them all [in reference to saunas], hotspots for queer contagion," "Health passports for homosexuals now!" (N2-ABC); "The faggots doing their thing," "Pure gay orgy," "Homo pox" (N2-EP); "They do things they shouldn't. It's not acceptable," "Grateful assholes, and the monkey is loose" (N3-MD); "Party, Maspalomas [city in the Canary Islands where Gay Pride event was held], men, rainbows? Nothing more to say" (N3-EP); "They excluded me for not being vaccinated, with a passport for medical reasons. Now you have to show your heterosexuality ID card, it's for your health," "Feminists spread the corona on 8M and the LGBTYs the monkey bug. The lobbies fucking with your health," "Pox Pride!!!!" (N4-MD).

A second set of comments comprises those opinions through which the social audience of these profiles tried to establish an analogy between the HIV/AIDS crisis and the emergence of the monkeypox outbreak outside Africa, once again targeting the gay population. This is, therefore, hate speech that, just as happened in the eighties, seeks to revive the twofold stigmatization of this group of people. Some examples of these comments are the following: "AIDS and now the monkey virus..." "Hahahaha and they say that the gays aren't the problem, I remind them that AIDS also started because of them," "First AIDS, now the monkey pox... Nature always punishes vice" (N2-ABC).

Finally, a third category of comments expressing hate speech toward gay men corresponds to expressions of contempt and hostility toward them simply because of their sexual orientation, which in some cases is defined as a disease. These are, therefore, posts in which the social audience on Twitter openly expresses disgust and contempt toward sexual minorities. Some of these comments are the following: "The only thing I like about Muslims is that they ban these nuisances [in reference to saunas catering to

gay men]," "Let's see if they can get cured and not get the pox while they're at it," "I applaud it [gay sauna closure]," "What sick people. This is what all this depravity creates," "And when will there be a vaccine for homosexuality?" (N2-ABC); "No comment. If they'd just sleep with whoever they want in private without these lurid exhibitions, maybe these things wouldn't happen. Freedom? Freedom" (N3-MD); "Not a party [in reference to Pride], a bunch of dirty pigs, looking for cocks" (N3-ABC); "The monkey pox thing is getting more and more repugnant. Information should not be censored, we want to know the origins of this virus. If it's really caused by the depravity and lechery of certain individuals, it would be alarming," "Antinatural," "A party of FAGGOTS" (N3-EP); "Good thing I'm not gay" (N5-EP). On occasions, hate speech was also expressed through memes, as the following image illustrates (Figure 6). It falls into this third category of comments and appeared as a response to two of the tweets analyzed (N2-ABC and N3-EP). In it, the following dialogue can be read between the two characters shown in the picture:

- Homosexuals should be confined for reasons of public health.
- Do you think that will do anything to stop the monkeypox?
- What monkeypox?

Figure 6: Meme expressing hate speech toward gay men

Source: Twitter.

Conclusions

The research undertaken has demonstrated the close connection that can be seen between disinformation, journalistic coverage from a conservative perspective, and hate speech against the LGBTIQ+ community on social networks, based on the study of how the Twitter audience reacted to posts about the emergence of the monkeypox outbreak in Spain. Disinformation and stigmatization went hand in hand for weeks, providing the perfect breeding ground for the generation of hate speech. As this study has shown, although the stigmatizing nature of the news was particularly pronounced in the content provided by a pseudo-media outlet specializing in the production of disinformation, homophobia also extended to the coverage of the events by conservative mainstream media organizations. The impact that websites that disseminate fake news and falsehoods can have on the discourse and style of hitherto trusted media outlets, with whom they are competing for the same audience on social networks, raises a number of questions that will need to be addressed in future research.

The media coverage, clearly homophobic, was partly a consequence of the fact that the first news about the outbreak came from an institutional source (the Community of Madrid) that needlessly identified a profile of those affected. This contributed to the stigmatization of gay men, in addition to generating a false sense of security among the rest of the population. Instead of highlighting the means or patterns of transmission of the virus, much of the information presented monkeypox as if it were essentially a disease affecting gay and bisexual men, emphasizing the places where transmission had occurred — an issue of zero relevance for raising awareness of the disease — and without contacting sources from LGBTIQ+ organizations (Canet, 2022).

This type of journalistic narrative, resembling a horror story, seeks to do more than merely inform. This is because its message, in addition to holding gay and bisexual men responsible, also has an implicit desire to scare them in order to discipline and instruct them so that they do not transgress

certain limits in their behavior: having sex with strangers, having sex with as many people as they want, or visiting certain places, among other red lines (Serrano, 2022). In order to put an end to this narrative, GLAAD and the CDC have created resources to combat anti-LGBTIQ+ stigma in monkeypox-related communications to make it clear that anyone can get MPV, and cisgender women have also contracted the virus. Thus, for example, they underscore that all stories that mention LGBTIQ+ people should include LGBTIQ+ voices, that the emphasis should be on how transmission of the virus is related to behaviors rather than to communities or identities, and that facts should be included that broaden public health alerts to reach the largest possible audience (GLAAD, 2022a).

Nevertheless, stigmatizing and disinformative discourse, generating hate messages, circulated freely on social networks during the first few weeks after the news broke. The internet was a powerful forum for both fake news (Czopek, 2022) and homophobic messages (Chan, 2022) during this initial period. Twitter failed to put a stop to such speech, despite the fact that many users utilized their accounts to spread dangerous hateful rhetoric against the LGBTIQ+ community, even insinuating that queer men had sexual relations with children by way of an explanation for the first cases of monkeypox detected in minors (Gallagher, 2022). This serious problem calls into question the platform itself in its ability to curb hate speech and also to moderate users' opinions — an area in which the role played by the media is also fundamental.

This study also highlights the possible relationship between clickbait (Gutiérrez, 2022) and the negative nature of the news shared on Twitter, and its capacity to generate engagement among users. Analyzing whether tweets that exhibit these characteristics not only favor the viralization of content but also audience reactions through opinions expressing hate is a matter which should be the subject of future research. Similarly, the results suggest the need to undertake further studies with the purpose of deepening the understanding of how hate speech interacts with a message's capacity for cultural resonance (the references to AIDS suggest that it has a

positive influence), the editorial policy of the media covering current events (Martínez Valerio, 2021), and the ideological and partisan polarization of the users who participate in the debate on social networks.

Funding

References

Albertini, P. (2012). Medios de comunicación. In L. G. Tin (Dir.), *Diccionario Akal de la homofobia* (pp. 338-344). Ediciones Akal.

Amnesty International (2021). *Amnesty International Report 2020/21. The state of the world's human rights*. Amnesty International. http://bit.ly/3X1XV8y

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info, 49*, 98-124. https://doi.org/10.7764/cdi.49.27817

Anti-Defamation League (2021). *Online hate and harassment. The American Experience 2021*. ADL's Center for Technology and Society. http://bit.ly/3UEPmPo

Arcila-Calderón, C., Sánchez-Holgado, P., Quintana-Moreno, C., Amores, J. J., & Blanco-Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: analysis of tweets with geolocation. *Comunicar, 30*(71), 21-34. https://doi.org/10.3916/C71-2022-02

Boone, R. (2022, June 13). Right-wing extremists amp up anti-LGBTQ rhetoric online. *Associated Press.* http://bit.ly/3hO8YCg

Canet, V. (2022, May 26). La verola del mico i l'homofòbia latent d'alguns mitjans. *Mèdia.cat.* http://bit.ly/3O8pDwg

Carratalá, A. (2023). Discursos mediáticos sobre el Covid-19 y sexualidad no normativa: viejos estigmas ante una nueva pandemia. In J. Yanes, M. Gabino, D. Fuentefría, & J. L. Zurita (Coords.), *El Periodismo, la Covid-19, la «gripe española» y las epidemias históricas* (pp. 183-200). Fragua.

Carratalá, A. (2019). La criminalización de la transmisión de VIH en los medios: Periodismo de sucesos entre el sensacionalismo y la homofobia. *Revista Española de Comunicación en Salud*, Suplemento 2, 38-51. https://doi.org/10.20318/recs.2019.4462

Carratalá, A. (2016). Press coverage of same-sex domestic violence cases in Spain. *Revista Latina de Comunicación Social*, *71*, 40-65. http://dx.doi.org/10.4185/RLCS-2016-1083en

Carratalá, A. & Herrero-Jiménez, B. (2019). La regulación contra el discurso de odio hacia el colectivo LGTBI en los medios: Análisis comparado de diez leyes autonómicas. *Revista de la Asociación Española de Investigación de la Comunicación*, *6*(12), 58-80. https://doi.org/10.24137/raeic.6.12.3

Chan, W. (2022, August 10). Rightwing media embraces Aids-era homophobiain monkeypox coverage. *The Guardian*. http://bit.ly/3V1eGi6

Clair, M. (2018). Stigma. In J. Michael Ryan (Ed.), *Core Concepts in Sociology* (pp. 318-321). Wiley-Blackwell.

Cruze Bell, J. S. (2006). Framing the AIDS Epidemic: From "homo" genous deviance to widespread panic. In L. Castañeda & S. Campbell (Eds.), *News and sexuality: Media portraits of diversity* (pp. 95-109). SAGE Publications.

Czopek, M. (2022, August 8). Image makes false claims about monkeypox and wasn't shared by BBC News or the WHO. *Poynter.* http://bit.ly/3EzFG32

da Silva, M. P. & da Silva, L. S. (2021). Disseminação de discursos de ódio em comentários de notícias: Uma análise a partir de notícias sobre o universo LGBT em cibermeios sul-mato-grossenses no Facebook. *Intercom-Revista Brasileira de Ciências da Comunicação*, *44*(2), 137-155. https://doi.org/10.1590/1809-5844202127

El País (2022, October 6). Delitos de odio. *El País*. http://bit.ly/3Tz6Zi5

Fejes, F. & Petrich, K. (1993). Invisibility, homophobia and heterosexism: Lesbians, gays and the media. *Critical Studies in Mass Communication*, *10*(4), 396-422. https://doi.org/10.1080/15295039309366878

FRA (2009). *Hate speech and hate crimes against LGBT persons*. European Union Agency for Fundamental Rigths. http://bit.ly/3O6RHju

Galop (2021). *What is online anti-LGBT+ hate speech and hate crime?*. Galop. http://bit.ly/3X2OPtY

Gallagher, A. (2022). *Monkeypox and 'Groomers': How Twitter facilitated a hate-riddled public health disinformation campaign*. Institute for Strategic Dialogue. http://bit.ly/3Ec78Ta

GLAAD (2022a). *Social media safety index*. Gay and lesbian alliance against defamation. http://bit.ly/3EyhHB9

GLAAD (2022b, June 29). Factsheet for reporters on monkeypox (MPV) and the LGBTQ Community. *Gay and lesbian alliance against defamation*. http://bit.ly/3UVcJ6Y

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Englewood Cliffs.

Gutiérrez, L. (2022, July 29). Viruela del mono: «Estigmatizar no cura, necesitamos respuestas». *Agencia Presentes*. http://bit.ly/3O5AwyZ

Hameleers, M., van der Meer, T., & Vliegenthart, R. (2022). Civilized truths, hateful lies? Incivility and hate speech in false information– evidence from fact-checked statements in the US. *Information, Communication & Society*, *25*(11), 1596-1613. https://doi.org/10.1080/1369118X.2021.1874038

ILGA (2021). *Annual review of the human rights situation of lesbian, gay, bisexual, trans and intersex people in Europe and Central Asia 2021*. International Lesbian, Gay, Bisexual, Trans and Intersex Association. http://bit.ly/3hHKYjZ

Keen, E. & Georgescu, M. (2020). *Bookmarks — A manual for combating hate speech online through human rights education*. Council of Europe. http://bit.ly/3E9PqQ7

Kowalewski, M. R. (1988). Double stigma and boundary maintenance: How gay men deal with AIDS. *Journal of Contemporary Ethnography, 17*(2), 211-228. https://doi.org/10.1177/08912418801702004

Kuhar, R. (2003). *Media representations of homosexuality. An analysis of the print media in Slovenia, 1970-2000*. Peace Institute. http://bit.ly/3EduGH8

Mangeot, P. (2012). SIDA. In L.-G. Tin (Dir.), *Diccionario Akal de la homofobia* (pp. 423-429). Ediciones Akal.

Martínez, R. (2016). *La cultura de la homofobia y cómo acabar con ella*. Editorial Egales.

Martínez Nicolás, M. A. (1994). La información periodística en la crisis del sida. Algunos temas de interés para la investigación comunicativa. *Anàlisi: quaderns de comunicación i cultura*, 16, 89-105. http://bit.ly/3Ad2a7t

Martínez Valerio, L. (2022). Mensajes de odio hacia la comunidad LGTBIQ+: Análisis de los perfiles de Instagram de la prensa española durante la "Semana del Orgullo". *Revista Latina de Comunicación Social*, *80*, 363-388. https://www.doi.org/10.4185/RLCS-2022-1749

Ministerio del Interior (2022). *Informe sobre la evolución de los delitos de odio en España - 2021*. Ministerio del Interior. http://bit.ly/3g5cgQQ

Palau-Sampio, D. & Carratalá, A. (2022). Injecting disinformation into public space: Pseudo-media and reality-altering narratives. *Profesional de la información*, *31*(3). https://doi.org/10.3145/epi.2022.may.12

Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *Sage Open*, *10*(4). https://doi.org/10.1177/2158244020973022

Rebollo Norberto, J., Generelo Lanaspa, J., & Assiego Cruz, V. (2018). *La cara oculta de la violencia hacia el colectivo LGTBI. Informe delitos de odio e incidentes discriminatorios al colectivo LGTBI 2018*. Observatorio redes contra el odio. https://bit.ly/3UFWafs

Rivera-Martín, B., Martínez de Bartolomé Rincón, I., & López López, P. J. (2022). Discurso de odio hacia las personas LGTBIQ+: Medios y audiencia social. *Revista Prisma Social*, *39*, 213-233. http://bit.ly/3ttjRvP

Rodríguez Martínez, E. N., & García Gavidia, N. (2006). Enfermedad y significación: Estigma y monstruosidad del VIH/SIDA. *Opción*, *22*(50), 9-28. http://bit.ly/3TxR7MY

Segura, M. S. (2021). Por una comunicación pública democrática. Alternativas a la desinformación y los discursos de odio. *Revista Sociedad*, *43*, 2-17. http://bit.ly/3UG9a50

Serrano, R. (2022, July 20). Asesinatos, virus y palizas: así difunden los medios terror en las personas LGTBI+. *Píkara Magazine*. http://bit.ly/3AhSatr

Tin, L.-G. (2012). Introducción. In L.-G. Tin (Dir.), *Diccionario Akal de la homofobia* (pp. 9-15). Ediciones Akal.

United Nations (2020). *Report of the Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, Victor Madrigal-Borloz Violence and discrimination based on sexual orientation and gender identity during the coronavirus disease (COVID-19) pandemic*. United Nations. http://bit.ly/3UIedlw

United Nations (2019). *UN strategy and plan of action on hate speech*. United Nations. http://bit.ly/3UWbeFJ

United Nations News (2022, May 22). Monkeypox: UNAIDS 'concerned' about stigmatizing language against LGTBI people. *United Nation News*. http://bit.ly/3g6sWr8

Ventura, R. (2019). *LGBT/Queer media studies: Aportaciones para su consolidación como campo de estudio* [Doctoral dissertation, Universitat Pompeu Fabra]. http://bit.ly/3AidE9D

Wardle, C. & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe*. http://bit.ly/3hL3szY

MAINSTREAMING THE FAR-RIGHT: DEPLATFORMING, CONTENT MODERATION AND FAR-RIGHT PRESENCE ON MAINSTREAM SOCIAL MEDIA SITES

Sarah Anne Dunne
/ University College Dublin, Ireland

Eugenia Siapera
/ University College Dublin, Ireland

"I'm Jewish, and there's a set of people who deny that the Holocaust happened. I find that deeply offensive. But at the end of the day, I don't believe that our platform should take that down because I think there are things that different people get wrong."

– Mark Zuckerberg.

In 2018, Facebook CEO Mark Zuckerberg controversially announced the right of Holocaust deniers to share their opinions online (Klein, 2018). The comment produced immediate outrage, garnered a paltry apology, and revealed the realities of online moderation and platform governance. While there are understandable and complex nuances behind platform governance policies and, particularly, moderation policies that can result in deplatforming, Zuckerberg's comment pointed to a deeper problem: that platform moderation was indelibly linked to the fact that social media platforms like Facebook remain, first and foremost, capitalist machines. As Zuckerberg went on to note: 'The principles we have on what we remove from [Facebook] are: if it's going to result in real harm, real physical harm,

or if you're attacking individuals, then that content shouldn't be on the platform' (Klein, 2018). Regardless of the ignorance behind such a comment and the undermining of online harm and digital violence, such a response (perhaps unsurprisingly) fails to account for the harm produced by groups often behind such content: white supremacists, antisemitic groups, and Nazi-apologists (among other extremist ideologues) produce this historical revisionism as a means of targeting Jewish communities and justifying Nazi propaganda and ideology (Cohen-Almagor, 2016). For companies like META, this points to an ideological blind-spot that refutes the realities of antisemitic hate speech (Guhl & Davey, 2020); an issue that traverses both antisemitism and social network platforms more broadly. Indeed, despite ongoing demands for policies and moderation that protects users from abuse and deplatforming actions taken by mainstream social media sites (MSMs), a growing body of work depicts the proliferation of extremist channels, including anti-LGBTQ+ and racist groups across these sites (see Gugl & Davey, 2020; Martiny & Lawrence, 2023; Rajan & Venkatraman, 2021). It has been further noted that these mainstream pages work as potential gateways into more radical content on alternative (or Alt-Tech) platforms (Mamié et al., 2021). This paper aims to highlight the growing use of MSMs by extremist and far-right groups and users often in spite or manipulation of the hate speech policies and user agreements. We intend to highlight how the continued use of these MSMs by extremist groups is reflective of both far-right ideologues' manipulation and side-stepping of hate speech policies and the vagaries of MSMs regulations, as indicated in Zuckerberg's earlier quoted statement.

The proliferation of far-right and extremist political discourse online is in no way a new phenomenon; far-right ideologues have utilised these networks since the circulation of Web 1.0 with websites such as Stormfront leading the cause and challenging the misunderstanding of such individuals as uneducated or unintelligent (Daniels, 2009), as is evident in their sophisticated avoidance of moderation tactics. Digital networks have always played a key role in the development of extremist and far-right communities. That

this growth has extended into the Irish context is, however, novel given the country's lack of significant far-right political party or political action prior to 2022/2023. It has only been in the last several years that a growing Irish far-right presence has been noted and connected to the Irish government's response to the Covid-10 pandemic. In particular, the disaffection and dissatisfaction of many citizens owing to extended lockdowns and restrictions became topics of ire and frustration across social media channels and, eventually, alt-tech platforms. Within this context, the Irish far-right flourished, reflecting the discontent of the larger population and proffering easy answers, exploiting the real disenfranchisement many felt via conspiracy theories that offered a quick return to the norm (e.g. that the virus was a hoax). Evidently, the radicalisation of Irish citizens via these groups did not stop with anti-vax narratives or covid-scepticism. Rather, far-right ideologues indoctrinated their new followers into more traditional and well-known far-right agendas which commonly vilified minority groups and challenged liberal or democratic politics. Climate denial, populist discourses and nationalism conjoined with anti-LGBTQ+ and anti-immigration narratives became common discussion points on both SMSs and alt-tech channels. While it is not within the scope of this work to cover all these topics, we intend to focus our analysis on how anti-LGBTQ+ discourse is shared on these sites. In doing so, we will provide a detailed content analysis of a video uploaded by Irish far-right journalist-cum-politician, Andy Heasman, in which he and a small group of supporters target an Irish bookstore for selling LGBTQ+ material.

Our analysis will focus on the use of language to consider how figures such as Heasman avoid the likelihood of their content being moderated or removed under META policies. Indeed, Heasman's avoidance of hateful language that targets LGBTQ+ people is replaced in favour of the language of protection and guardianship, with the symbolic figure of the child being central to their argument, hence avoiding META's regulation policies. In order to create a rounded discussion of these policies – and the context in which Heasman and other far-right ideologues and extremists are managing to

share their content despite hate speech and anti-extremism regulations – we will begin this paper with a historical overview of platform governance policies and the historical contexts which led to the ongoing moderating and deplatforming of far-right. It is our contention that these policies have informed how the far-right are currently sharing their political messages on MSMs, and, furthermore, that such posts may act as gateways into more radical and extremist positions.

From early bans to the great deplatforming

Deplatforming, or the permanent banning of an individual, group, or organisation from a platform, has long been a contentious topic that the far-right have frequently framed as, at best, attempted censorship and, at worst, tech companies bowing to liberal pressure. Controversies ranging from the effectiveness of deplatforming in disengaging radicalisation of users to whether free speech trumps hate speech are common reservations from both sides of the argument. As far as mitigating the reach of extremist discourse and the potential for radicalisation, the effects of deplatforming cannot be undermined. In 2016, when alt-right darling and Breitbart news editor Milo Yiannopoulos was banned from Twitter[1] for his role in organising a targeted hate campaign against comedian/actress Leslie Jones, few could foretell the influence such a move would have over his career. While Yiannopoulos may have initially celebrated the decision (Romano, 2016), he later claimed that removal from MSMs effectively ended his career and rendered him broke (Klein, 2020).

Yiannopoulos' removal was not strictly unprecedented given his frequent suspension owing to various policy violations and harassment of other users. Significantly, the use of temporary suspensions represents another facet of the deplatforming strategy: other methods including content

---

1. Yiannopoulos would later be banned from other MSMs, including Youtube and Facebook.

Mainstreaming the far-right: deplatforming, content moderation
and far-right presence on mainstream social media sites

removal, warning or strike systems, demonetization[2] and shadow-banning[3] are further implemented by various platforms as a means of encouraging stronger engagement with the user policies. Regardless of these other methods, Yiannopoulos' removal surprised many even beyond his loyal followers and fellow trolls; not least because such big tech platforms had long been accused of purposefully maintaining, propagating, and profiteering off such content (Kirk & Schill, 2024). Twitter, at the time, responded with a commitment to improve their moderation and removal policies having also received negative backlash following their handling of Gamergate: 'We have been in the process of reviewing our hateful conduct policy to prohibit additional types of behaviour and allow more types of reporting, with the goal of reducing the burden on the person being targeted' (Romano, 2016).[4]

In 2018, Infowars conspiracist and radio-show host Alex Jones faced similar repercussions as his channel was removed from Facebook, YouTube and streaming services Spotify and Apply, with Twitter later following suit. As with Yiannopoulos, Jones had faced a series of repercussions and warnings for sharing harmful content on his podcasts and channels, and for the type of conspiracies and disinformation his channels traded in, particularly in rebranding national tragedies as hoaxes. Jones' frequent reiteration of the belief that the Sandy Hook Elementary school shooting – which tragically took the lives of 26 people, 20 being children in December 2012 – was a manufactured hoax perpetrated by a liberal government aiming to restrict gun laws, identifying it as a 'false flag' operation, referring to the grieving parents and survivors as 'crisis actors' (Sellnow et al., 2019: 131). While the MSMs which banned Jones did not directly cite the reason for removal,

2. Demonetization, as a strategy of platform governance, involves removing the opportunity for verified users to monetize their channels, effectively removing their ability to make money via the content they provide and/or the ads and sponsorship they endorse.
3. Shadow-banning involves the partial blocking of a user's content from reaching their followers and other audiences. While the content remains available on the user's page, it effectively does not appear on follower's feeds or updates, causing a lack of engagement. This tool often means content searches will often exclude a shadow-banned account's posts also.
4. The harassment campaign known as Gamergate involved an orchestrated attack primarily against women in the gaming community. While many claimed the campaign was about ethics in the larger gaming industry, Gamergate has been noted for a series of virulent and elongated online attacks against women gamers, designers and content creators, including game programmer Zoe Quinn and gaming vlogger Anita Sarkeesian.

references to policy violation and hate speech along with violent discourse were common across the board (Coaston, 2018). Reflecting on Zuckerberg's printed quote and the noted polarisation he espouses between disinformation and hate speech, it is worth citing that Facebook clearly differentiated from others in locating hate speech and *not* disinformation, alarmingly dividing two highly intertwined narrative devices: 'While much of the discussion around Infowars has been related to false news, which is a serious issue that we are working to address by demoting links marked wrong by fact checkers and suggesting additional content, none of the violations that spurred today's removals were related to this' (Coaston, 2018). Jones, again mirroring the effects of deplatforming on Yiannopoulos, has similarly claimed bankruptcy and, moreover, is facing charges as a result of defaming the Sandy Hook families (Robertson, 2022). Evidently, both Yiannopoulos and Jones' removal from MSMs has been linked to a noted regression in their celebrity status, limiting their overall reach to other users and effected their wealth.

These early examples of mainstream deplatforming provide evidence that removal of persons from MSMs – and removal of the extremist content therein – is indeed a successful method for undoing and undermining far-right extremism and hate speech. Indeed, Jhaver, Boylston, Yang and Bruckman's analysis of deplatforming as a tool found that the removal of far-right actors (in this case, Yiannopoulos, Jones, and comedian Owen Benjamin) had resulted in not only a drop in posting activity of their supporters but also reduced the 'overall toxicity levels of supporters of each influencer' (2021: 4). The findings indicated the 'efficacy of deplatforming offensive influencers to counteract offensive speech in online communities' (Jhaver, 2021: 4). Other researchers produced similar results, finding deplatforming to be an effective tool of platform governance for mainstream sites (Rauchfleisch & Kaiser, 2021; Rogers, 2020). Reddit's decision to remove fat-shaming and racist subreddits in 2015, for example, was determined a success as offensive and hate-fuelled content apparently decreased (Chandrasekharan et al., 2017; Saleem & Ruths, 2018), while others have

'found that deplatforming significantly reduce[s] the popularity of many anti-social ideas associated with influencers' such as Yiannopoulos and Jones (Jhaver et al., 2021). Deplatforming has additionally been found to mitigate the spread of disinformation somewhat, though the sharing of content often continues via new account or new sharing streams and at times is more a disruption in the content stream than a deterrent or undoing (Bruns et al., 2021; Innes & Innes, 2023).

The Covid-19 years witnessed a series of deplatforming's for influencers who peddled in misleading and harmful health narratives; disinformation was rife throughout this period, with MSMs introducing visible fact checking methods and issuing warnings on posts related to the pandemic. With the pandemic and these sweeping measures came new, alternative platforms that offered less stringent moderation policies and greater leniency to its users. Alt-tech platforms opened their arms to new users and known influences such as Yiannopoulos and Jones (who invited their fanbases to follow their new channels) and profited. YouTube replica Bitchute and Twitter-esque platform Gab both boasted increased users and traffic owing to deplatforming measures (Rauchfleisch & Kaiser, 2021) pointing to the greater issue: 'When deplatformed social media celebrities migrate to alternative platforms, these sites are given a boost through media attention and increases in user counts' (Rogers, 2020: 214). The migration of deplatformed users and their followers to such platforms could boast further negative effects, rendering it more difficult to police and navigate extremist threats. In some cases, it has been highlighted that deplatforming, while restricting the reach and audience of extremist content, has rendered users more hostile and active (Ali et al., 2021) and may contribute to the production of deeper radicalisation (Buntain et al., 2023: 2; Urman & Katz, 2022: 908). The disadvantages of deplatforming are evident, and there remains serious debate on how and why MSMs utilise it.

In this regard, there are legitimate concerns to how deplatforming is enacted and there are debates that more resemble strawmen. For the latter group, deplatforming is often presented as a form of censorship that

suppresses free speech, thereby threatening core human rights and freedoms, as Alex Jones compared his deplatforming as akin to 'communist style censorship' (Rauchfleisch & Kaiser, 2021: 5). The free speech debate further consolidates the conspiratorial belief that governments and tech companies either highly influenced or outrightly controlled by liberal elites (Alfino, 2023: 11; Mulhall, 2019). According to communications researcher Prashanth Bhat, the liberal bias argument is not novel but rather has been an ongoing narrative device of conservative groups and is being 're-weaponized by the American Right to challenge the credibility of corporate social platforms' (2022: 109). The right and responsibility of such corporations to protect themselves legally and to uphold their user agreements and the values espoused (see van Dijick et al., 2023) are largely disregarded within this narrative frame. Yiannopoulos replicated these exact discourses in his lamentation of Twitter becoming a 'safe space' (Romano, 2016) at the expense of open debate and free speech. This use of language invalidates liberal politics and ideologies as potentially coddling and certainly too "politically correct" as a result of practices that include providing trigger/content warnings or offering safe spaces to marginalised groups. As for labelling these discourses as strawmen, there is significant irony behind the far-rights considerable fight for free speech when it suits their agenda. Indeed, the far-right continuously undermine the rights of minority groups to be heard; as this paper will go on to depict, currently many far-right groups are engaging in library and bookstore protests which disputes the availability LGBTQ+ content. As Adrian Rauchfleisch and Jonas Kaiser have noted: 'calling deplatforming censorship, then, is often more a rhetorical weapon to avoid acknowledging the4 spread of disinformation, racism, or other extremist speech, than an accurate and honest analysis' (2021: 6).

Scholars, by contrast, query the fairness and bias of these procedures, expressing concern over how deplatforming and other forms of moderation are operationalized and managed. The moderation and removal of content and users from MSMs often involves policies and actions that are non-transparent and difficult to challenge. The innate human bias involved

in both manual content moderation and AI or automated content moderation make the likelihood of mis-contextualisation (relating to language use or other circumstantial instances), errors and biased decision-making; further involved and prone to biases are user reports, content flags and other automated tools (Díaz & Hecht-Felella, 2021). Indeed, concerns abound in this regard:

> 'A growing body of scholarship has documented the multip0le challenged with commercial content moderation as enacted by platforms today, ranging from labour concerns (about the tax concerns) about conditions and mental health challenges faced by moderators, many of whom are outsourced contractors in the Global South); democratic legitimacy concerns)about global speech rules being set by a relatively homogeneous group of Silicon Valley elites); and process concerns about the overall lack of transparency' (Gorwa et al., 2020: 2).

Human rights advocates and academics, such as the Santa Clara Principles have argued the necessity of transparency and accountability in moderation policies and platform governance from big tech companies; and while many of these MSMs now offer regular transparency reports and appeals procedures, the processes behind deplatforming, moderation and the management of data more broadly remains opaque and, often, selective. The widespread censoring, for example, of pro-Palestinian stories from META's platforms has been noted by the Human Rights Watch despite much of the data being non-violent and yet was 'unduly supressed' in a 'systemic and global' manner (Brown & Younes, 2023). A majority of MSMs are governed by for-profit enterprises and, as such, prioritise capital accumulation.

This fact has not escaped broader public attention, with headlines such as the following emerging: 'social media companies prioritising profit over harmful content: Senate report' (Barr, 2022). There is a near impossibility in proving such statements, given the convoluted and cryptic characteristics of content moderation and data management within these companies. Indeed, potent questions abound when powerful and influential figures

seem to post with impunity while others are subject to strict regulation (Díaz & Hecht-Felella, 2021: 15-16). Thus, the argument remains: 'On social and digital media, algorithms reinforce and amplify outrage and extremity. Both legacy as well as digital and social media have mastered the art of monetizing anger, paranoia, and distrust' (Kirk & Schill, 2024: 6).

Public intervention and global events, as much as users' actions, can impact wide scale deplatforming, as was witnessed with the covid-19 pandemic and the spread of disinformation during this time. Indeed, the removal of Yiannopoulos and Jones from MSMs in 2018 has been linked to growing public outrage over far-right violence that had occurred at a demonstration in Charlottesville in late 2017. The Unite Right Rally unified many cohorts of the alt- and far-right including Neo-Nazis, nationalists, and white supremacists, one of whom would deliberately drove a vehicle into a crowd of counter-protestors, injuring 35 and killing Heather Hayes in the process. Trump's election, growing far-right popularity across the West and other political motivations heralded new opportunities for these right-leaning groups to (re)legitimise their political efforts, which the Unite the Right rally capitalised on. MSMs played a major role promoting the rally (Donoval et al., 2019); a reality which many of these corporations took great pains to distance themselves from after the violence of the event. In an apparent attempt to mitigate claims that they had served as propagators (and profiters) of the rally during its organisation and promotion, many MSMs responded with serious moderation upheavals (Donovan et al., 2019). Referred to as the 'purge' (Rauchfleisch & Kaiser, 2021: 9), the deplatforming of high-profile, far-right ideologies such as Jones and Yiannopoulos, was perhaps more politically motivated to protect MSM corporations than community standards.

Rita Kirk and Dan Schill have produced similar results in their research of the Jan. 6 Capitol siege, during which staunch Trump supporters stormed the Capitol building in Washington, adamant that election fraud had secured Trump's loss. MSMs – which were 'foundational before, during, and after the attack' (Kirk & Schill, 2024: 6) – were again in the hot seat and held responsible for the disinformation which circulated regarding electoral

fraud and, moreover, for the consistent display of leniency granted to Trump whose messages frequently espoused violent and hate-filled content along with gratuitous and dangerous disinformation (Díaz & Hecht-Felella, 2021: 16). In what has since become termed "the Great Deplatforming," Trump, and other key figures involved in the Capitol siege and in touting electoral fraud disinformation, were permanently banned from several MSMs. In conservative fashion, Trump would be-cry the loss of free speech and claim that MSMs were working at the 'behest of Democrats and the radical-left' (Bhat, 2022: 108). Many users would follow their far-right ideologues to the alt-tech sites that were decidedly unlike mainstream platforms (Bhat, 2022: 111), many of whom would relocate their views to more lenient alt-tech sites. This, along with promises of user anonymity, end-to-end encryption, and privacy, renders alt-tech sites as virtual havens for extreme content creators and the far right to share their politics and ideologies, build communities and movements, and establish legitimacy even with regard to the disinformation and hate-fuelled content they share.

Hate speech and anti-LGBTQ+ narratives

While alt-tech sites provide ideal spaces for the growth of far-right political groups, this does not negate the continued significance of MSMs for them too. As stated, alt-tech sites provide ample opportunities for growth but not the reach or audience present on more mainstream spaces. The importance of these more populated virtual spheres as potential gateways into more radical and extremist content remains and they are still sought out as spaces to share far-right content as such, even despite the risk of content removal, as guarantors of a mainstream presence. Indeed, as this paper goes on to argue, many far-right users have adopted savvy techniques to sidestep content removal and deplatforming, locating the nuance in MSMs hate policies and user agreements in a way that makes the content acceptable. As has been highlighted, the nuances and complexities of language, context and situation can render the concept of hate speech as questionable; even the use of slurs is debatable when key target groups are reclaiming them

as a means of power (Gorwa et al., 2020: 10). We need only refer back to the opening quotation from Zuckerberg to recognise that for MSMs, content moderation of hate speech is a complex process.

Myriad definitions and vague policies further exacerbate such complexities, lending little in the way of direction for moderations or those flaggers who report on content they deem harmful. Given this, it is perhaps redundant to state that hate speech lacks a singular definition. Rather, we can consider hate speech as referring to discourse which dehumanises or discriminately targets historically othered and minority groups; as incitements to violence against these groups; and as means of undermining the agency and self-determination of said groups (see Daermstadt et al., 2019; Guhl & Davey, 2020; Saresma et al., 2020). Furthermore, as Zuckerberg's statements fail to recognise, hate speech and disinformation as narrative tactics are often intertwined, fabricating events and discourses to undermine the rights and existence of minority and target groups.

All MSMs prohibit hate speech in some form, as Tarleton Gillespie argues it is 'a safe position politically' (2018: 59), and often reiterates the very community values that these sites lay claim to. Facebook/META's mission statement has always been 'to build community and bring the world closer together' while one of META's key principles is to 'keep people safe and protect privacy,' highlights their apparent commitment to their community's wellbeing (META). META's defines hate speech as 'direct attacks against people – rather than concepts or institutions – on the basis of what we call protected characteristics' (META). The platform further recognises dehumanising speech, stereotyping, expressions of inferiority or segregation as acts of hate speech. Such abstract definitions provide fertile ground for manipulation and circumvention; far-right actors such as Heasman are especially tactical in their avoidance of egregious language or slurs and will reference targeted groups in roundabout ways, focusing their attention on those that they reconceive of as being most vulnerable and at risk as a means of undermining and discrediting the group they are truly targeting. These tactics provide opportunities for the far-right to further deny their

intolerances and discriminatory politics in the guise of protecting another vulnerable group. Prime examples of this include references to the Irish housing crisis and homelessness when protesting migrant accommodation and asylum seekers rights and, as will be discussed, a focus on children and childhood to contest LGBTQ+ visibility.

Such tactical measures are not only elusive of META's slippery policies, they also encourage further public engagement and sympathy via such manipulative discourse. As has been highlighted, META is keen to outline the nuances apparent when enforcing their own policies, as Zuckerberg's irresponsible quote suggests. More succinctly, the reality of content moderation as a tool is challengeable when context cues and acceptable language are elusive as they tend to be on social media. Indeed, while the use of slurs are not permitted according to META's policies, their reclamation as an act of power and transgression by certain minority groups provides ample ground for reconsidering the effectiveness of content moderation more generally. And indeed, the reliance on and use of these methods remains debateable with Zuckerberg claiming before US congress in 2020 that Facebook hate speech moderation was at 94 per cent and META's transparency reports locating actioned content as consistently between 80 and 90 per cent; whereas leaked documents seem to place the number at a more meagre 3 to 5 per cent (Giansiracusa, 2021). Meanwhile, international NGO Global Witness sought to test Facebook's tolerance toward hate speech, introducing ads to the platform which targeted and dehumanised marginalised groups and incited violence against them. The ads varied in relation to location and content but were explicitly hate-fuelled and were sent for publication to Facebook, TikTok and YouTube; of the three, Facebook accepted the majority (Peck, 2023). Such findings query not only the actual statistics but the very efficacy of the policies which META is governed under. Between such slippery definitions and potentially false statistics, the question of how hate speech proliferates on such platforms is a pertinent one, particularly as the far-right community grows across them.

Homophobia is well documented within these far-right communities, with anti-LGBTQ+ narrative constituting a primary aspect of their political ideologies (Mudde, 2019). In particular, recent years have witnessed staunch and highly discriminatory rhetoric and actions against trans and queer communities. For the far-right, these groups represent real threats to the heteronormative order and the traditional family structure (Leidig, 2023: 75). Fears of indoctrination and grooming are strongly articulated across far-right digital channels (Leidig, 2023: 90); often, these narratives are connected to the concept of "gender ideology," and feminist/queer theory which extends gender beyond binary restrictors and opens sexuality beyond the heteronormative. The perceived threat which this poses to the natural order and nuclear family thus becomes a central concern for the far-right. Within these narrative frames, the figure of the child is often central and symbolically placed as being in need of protection and support (Slothouber, 2020: 93). Crucially, such rhetoric often overlooks the needs of trans and queer youth who may be seeking such supports and resources.

In the eyes of many far-right and conservative groups, LGBTQ+ youth do not exist but are rather groomed or indoctrinated via various liberal ideologies. One broad frame often used in this is related to what they call "gender ideology," a perceived liberal agenda that the far-right argues is undoing the biological order of sex and sexuality and which threatens natural gender norms. The far-right deploys these arguments as a 'political instrument' that delegitimise liberal and feminist politics (Corredor, 2019: 616): 'The use of the term *gender ideology* functions discursively to bring together different forms of right-wing mobilization, united against progressive demands for gender equality and against tolerance for sexual diversity' (Leidig, 2023: 80). It is crucial to note that these discourses do not exist in a vacuum, rather with the growth of these narratives online there have emerged simultaneous upswings in real world hate crimes and acts of violence (Godzisz & Vigiani, 2019).

In early 2023, the Irish police force, An Garda Síochána, released a report noting a 29 per cent rise in hate related incidents, with LGBTQ+ people

marked as the second most targeted group following racial minorities (Fanucci, 2023). This statistic and the marked rise in far-right political agitation in Ireland cannot be undermined. Indeed, over 2023 there was a marked rise in far-right demonstrations against LGBTQ+ resources and activities across Ireland. This has included the intimidation of groups organising and supporting trans events (Linehan, 2023) and a rally organised by TERF activist, Posie Parker, and anti-trans group, Let Women Speak, in Dublin city centre[5] (McGreevy, 2023). Libraries and bookstores have been primary targets for these groups throughout 2023, with demonstrations protesting the availability of LGBTQ+ material for minors and young adults (Fitzgerald, 2023). Such protests, again, overlook the necessity of such material and resources for LGBTQ+ youth while hypocritically insisting that their agenda is to safeguard Irish children. Simultaneously, their focus on texts which prioritise queer sexual health acts to further stigmatise LGBTQ+ sexuality and sexual acts more broadly.

Protesting LGBTQ+ material

The aim of this analysis is twofold: firstly, to provide a critical analysis of the anti-LGBTQ+ rhetoric used by far-right ideologues on MSMs and secondly, to consider how this rhetoric often circumnavigates MSM hate speech policies and community/user agreements. In order to do so, we provide a detailed content analysis of the dialogue taken from a Facebook video post uploaded to far-right activist Andy Heasman's public page. The post depicts a demonstration at Irish bookstore chain, Dubray, where Heasman and a supporting cohort of protestors, object to the availability and sale of LGBTQ+ sexual health and lifestyle material to minors. Heasman, who takes the lead in the filmed protest, has earned himself a name as a far-right agitator across Ireland, and reached notoriety during the Covid-19 pandemic with the circulation of anti-vaccine disinformation and his arrest for purposefully disregarding the mask mandate on public transport. While advertising himself as a freelance Irish journalist and declaring intentions

5. Importantly, a counter-protest simultaneously took place celebrating queer and trans lives.

to run as an independent candidate in the 2024 Irish elections, Heasman is widely known for his involvement in far-right protests, in particular he has championed himself as a protector of children via ongoing library and bookstore protests (Fitzgerald, 2023). Indeed, Heasman regularly posts about these protests to his multiple social media accounts. The Facebook video we intend to analyse remains active on Heasman's public Facebook profile: it was uploaded 25 June 2023 and runs for a total of 27 minutes and 2 seconds, during which Heasman and his accomplices attempt to debate the indoctrination of Irish children by Dubray staff and the LGBTQ+ material they stock. We will provide both a descriptive content analysis of the transcript and a visual analysis of the video to provide a detailed examination of the encounter and compare the results with Facebook policy in our analysis.

As stated, Heasman's video takes place in Co. Cork bookshop, Dubray, and takes aim at the LGBTQ+ resources available through their premises. Heasman, behind the camera, is joined by fellow far-right agitator Ross Lahive and several others, two members of Dubray staff and, later, intervening Gardai (who Heasman himself requested under the belief that a crime is taking place via the sale of such materials to minors). The books in question, *This Book is Gay* and *What's the T?* both by Juno Dawson (a transgender woman), are best described as introductions and guides on different aspects of queer identity aimed at teens and young adults and include information on sexual acts and safe sex. *What's the T?* focuses primarily on questions that transgender and non-binary teens and young adults may be grappling with, while Dawson herself has referred to *This Book is Gay* as 'a manual to all areas of life as an LGBT person' (Dawson, 2014). Heasman and his supporters, however, are quick in their video to identify both works as 'pornography' (2023). Their primary argument insists that as such, it is being illegally marketed and sold to children according to the Irish Child Protection Act 2015. Also known as Children First, this legislature highlights that sexual abuse against a child includes 'wilful exposure of the child to pornography' (Children First Act 2015).

Heasman and his support utilise this line of argument throughout, contending that Dubray and its staff are engaging in illegal activities via the sale of such material to minors; at the presence of Gardaí, Heasman and Lahive become insistent on this fact and call for arrests of staff or removal of the materials. Two points of note here: both books are advertised for teens (not children as the videos falsely and frequently states) and are visibly marked as having 'mature content' via a sticker applied by Dubray (Heasman, 2023). Evidently, the mature nature of the content is recognised and regarded seriously, as is the case with many teen and YA books available within such stores and which were not taken to task by Heasman or his supporters. Similarly questionable is why these protests do not challenge the availability of texts which detail sexual health and well-being for heteronormative youths. While this point alone does suggest the hypocritical and discriminatory nature of such protests, it remains notable throughout this video that references to LGBTQ+ identities are almost non-existent (apart from a transphobic instance at the end of the video). Rather, Heasman and Lahive carefully redirect their language away from the issue being LGBTQ+ identities or materials and concentrate more on the potential indoctrination or sexualisation of children that could occur as a result of such material. This discourse itself relates to an entire history of myths and prejudices that conflated homosexuality with paedophilia, historically misidentifying queer men as threats to children (see Bennett et al., 2024). Such rape myths pathologized homosexuality and queerness as deviant and perverse (Bennett et al., 2024: 1); an iteration which is implicit in the arguments made by Heasman, Lahive and his supporters.

The video begins abruptly with Heasman and Lahive already mid-debate with the Dubray staff, querying the age group the books are sold to, specifically asking whether a bookseller would hypothetically sell the material to a ten-year-old. Both declare that the store and its workers have a 'duty of care' to its consumers and, primarily, the hypothetical children purchasing such material (2023). The protestors are quick to claim that the material actively 'grooms' and 'indoctrinates' children, holds the booksellers accountable for

selling this "porn" illegally to minors, and highlights their moral ground by painting themselves as guardians of children. They are 'standing up for defenceless children' and express concerns that the material is 'damaging to children's health' (Heasman, 2023). Implicit here, of course, is the false equivalency that historically has linked homosexuality and illness/disease which was heightened during the AIDs epidemic (see Sontag, 1989). Such statements and beliefs continue to produce prejudicial feelings and acts against the queer community, displaying further disregard for children and teens who are likely to identify beyond the heteronormative. And indeed, despite claiming their latent guardianship of Ireland's children, at no point does the likelihood of queer youth necessitating access to such material considered. Rather, it is ostensibly suggested via their argumentation that children and teens do not identify as LGBTQ+ so much as they are indoctrinated into it. The implicit homophobia behind these arguments becomes clear again when Lahive and Heasman condemn what could be considered the crasser information the text proffers in sections that detail safe anal douching and oral sex. Lahive and Heasman both reference the texts as 'despicable' and 'filth' at this and other points of the post (Heasman, 2023). The reactions of disgust and moral outrage expressed in this instance not only stigmatises queer desire and sex but threatens safe and easy access to one of the few LGBTQ+ sexual health resources available to queer youth.

When visible on video, several of the protestors are seen to be wearing 'education not indoctrination' t-shirts, a popular slogan of the far-right which highlights the concern that the teaching of inclusive gender identity and sex-ed in schools (and other instructional institutions such as libraries) is actively grooming Irish youth (O'Connor, 2023). As aforementioned, what is clear via this ideology is the symbolic use of the child as a marker of the need for anti-LGBTQ+ political action. The existence of such groups and the resources they require are framed as threats to children's health and well-being. Van Slothouber, in their work on mainstream media and stories of detransitioning, highlights the significance of the child as a symbol in anti-trans (and anti-LGBTQ+) politics. Following queer theorist Lee Edelman's

Mainstreaming the far-right: deplatforming, content moderation and far-right presence on mainstream social media sites

theory in *No Future*, 'the Child remains a figurative child, "not to be confused with the lived experiences of any historical child" (Edelman, 2004: 11) [...] This ideological discourse works to deny citizens of their rights in the here and now, instead holding out for this future Child' (2020: 93).

In this respect, we can consider the means by which the discourses Heasman, Lahive and his cohort utter as often undoing the existence of queer children or teens as a matter of fact. Indeed, when Heasman states that children 'shouldn't be able to stumble across filth like this' (Heasman, 2023) there is little to no regard for children or teens actively searching for such resources; in fact, for these groups, such children are indoctrinated into seeking this material rather than seeking it out at their own initiative and will. This, in itself, points to the manner in which the group manages to avoid almost completely any discussion of LGBTQ+ people in this work. Keeping in line with META's hate speech policies, Heasman and his cohort avoid directly referencing any protected or vulnerable groups directly, in fact they manage to mostly avoid any utterance of the words gay, lesbian, or queer. Rather, they are able, through the symbol of the child, to focus their attention on the texts and direct the problems associated with these groups to the health and well-being of that hypothetical being.

The only instance in which queer identities are specifically mentioned and targeted occurs towards the end of the video when, in a bid to depict the varied moral transgressions of the bookstore, Lahive takes aim at trans actor Elliot Page's memoir, *PageBoy*. The book, wherein Page recounts his coming out as both queer and trans, and the book-jacket depicts Page's transformation following top-surgery and hormone replacement therapy. In the video both Heasman and Lahive direct their ire towards Page depicting their transphobia: both deadnaming the actor and repeatedly using the wrong pronouns, flippantly noting:

> Heasman: She [Page] was gorgeous.
> Lahive: Gorgeous, you know?
> Heasman: Look at the state of her now.

This discourse points to the inherent misogyny and sexism rife within anti-trans and far-right politics, where women's worth becomes linked to beauty and reproduction primarily. A later altercation in the video further depicts said beliefs, a customer purchasing a queer text is approached by Heasman, who informs him that the store 'sells porn to kids.' The customer, themselves purchasing a queer themed book, calls Heasman a 'Nazi' leaving Heasman to condescendingly ponders if the man is a 'predator' simply for purchasing a queer-themed text, again reiterating these links (Heasman, 2023). The sharing of such disinformation is a key aspect of these groups, and often allows them to push specific agendas. As noted at the beginning of this work via Zuckerberg's quote, disinformation is not necessarily of great concern for META: indeed, the right of Facebook's users to express different ideas and opinions is protected in their policies just as people are protected from hate speech and discrimination on them. Disinformation, in this case, exaggerates the content of Dawson's texts and, in doing so, poses them as sexually licentious and too advanced for their age group.

This is evident again in Lahive and Heasman's recontextualising of the sexual material available within the book. Despite its matter-of-fact style and that it comes from a trustworthy source (rather than queer youth having to seek out such information from strangers on the internet, for example) and offers salient advice on safe sex, STIs and queer stereotypes, Lahive and Heasman demonise these texts via their expressions of moral outrage and repulsion. Another protester falsely states that the author explicitly tells their readers to join queer dating app Grindr, a platform that is known for its proclivity and hook-up culture. Similarly, when discussing the issue with Gardaí, Heasman also incorrectly states that the books are 'telling 11-year-olds to go onto these apps [Grindr] where predators can get them' (Heasman, 2023). While the book does contain information of popular queer sites and dating apps, particularly Grindr, it is balanced in its objectives pointing to the positives and negatives of such sites, as well as providing advice on how to navigate them and engage with other users. Moreover, Dawson explicitly states in the print that such sites are for 18-year-olds plus with emphasis

(2021). The larger philosophical debate that remains here is that while it may seem jarring that this content is aimed at teens, the text remains the more tasteful and secure option for queer teens who may otherwise find themselves out of their depths or would have to circumnavigate the disinformation, homophobia, and extremist content (porn included) of the internet to find such information.

The rhetoric in this video, while tellingly anti-LGBTQ+ in its aim, is tame in its rhetoric. Barring the transphobic statements relating to Page, there is little in the way of a verbal, targeted attack on the queer community. This restructuring of anti-LGBTQ politics into a more palatable discourse has become an essential element in the (re)legitimation of far-right political groups in recent years. Indeed, across platforms, far-right ideologues are cautious of the language and ideologies they tout, particularly figures such as Heasman and Lahive who have advocated for far-right political candidates and have considered or are entering the electoral race themselves.[6] Tellingly, however, this does not extend to the followers and audience of posts who are free to express themselves more egregiously. While many avoid explicit homophobic and transphobic remarks on MSMs, the use of alt-tech platforms – which has grown exponentially in recent years – has encouraged and permitted the use of hate speech and anti-LGBTQ+ discourse. On Heasman's public Telegram channel, for example, his followers offer a multitude of responses that range from messages of support to a use of slurs and reaffirmations of queer stereotypes and rape myths that connect trans people to groomers, predators and paedophiles, and calls for a regression of LGBTQ+ right such as gender affirming health care for trans people. The fact that such rhetoric is shared without impunity on these platforms, and that the administrators like Heasman do not act to remove such posts or remonstrate the original poster, demonstrates their very acceptance. As with most echo-chambers, those who denounce or challenge

6. Heasman had announced his intention to stand as an independent candidate in late 2023 but withdrew his bid online in January 2024. Lahive is still considering running for Irish election as an independent candidate in 2024.

these discourses are marked as Other; as with the customer in Dubray who dared denounce Heasman's politics, these denouncers are posited similarly as dangers and threats.

What is evident via this analysis is that while the rhetoric has changed, the ideology has not; these far-right ideologies are rather the wolf in sheep's clothing on MSM platforms, maintaining their regressive stance to undermine LGBTQ+ gains and challenge them at any opportunity via new discursive techniques and tactics. In their deferral to the symbol of the innocent child, the Irish far-right is building a more acceptable route to homo- and transphobia; those who are reiterating anti-LGBTQ+ messages can defend themselves not as bigots or discriminators but as merely concerned about Ireland's future and the next generation. These strategies have been essential to the development of the Irish far-right and the ubiquity with which their messages are spreading to the broader public. They have further permitted the far-right to remain present across MSM platforms despite ongoing moderation and deplatforming efforts. What remains evident is that while hate speech policies remain vague and undetermined in their moderation of extremist and discriminatory content, the far-right will continue to maintain a mainstream presence and employ tactical means such as those shown to ensure the circulation of their narratives.

Acknowledgement

References

Alfino, M. (2023). Free speech issues in the regulation of social media platforms. *Journal of Information Ethics*, 32(1), 10-14.

Ali, S., Saeed, M. H., Aldreabi, E., Blackburn, J., De Cristofaro, E., Zannettou, S., & Stringhini, G. (2021). Understanding the effect of deplatforming on social networks. *Proceedings of the 13th ACM Web Science Conference 2021*, 187-195.

Mainstreaming the far-right: deplatforming, content moderation
and far-right presence on mainstream social media sites

Barr, L. (2022, 16 Nov). Social media companies prioritizing profit over harmful content: Senate report'. *ABCNews.* https://abcnews.go.com/Politics/social-media-companies-prioritizing-profit-harmful-content-senate/story?id=93360057

Bennett, M., Harris, D. A., Richards, K., & Shields, R. (2024). 'Any male victim?' The legacy of conflating paedophilia and homosexuality and the differential seriousness of sexual abuse by victim gender. *Psychiatry, Psychology and Law,* 1-15.

Bhat, P. (2022). Counter-net of tomorrow?: Right-wing responses to deplatforming Trump. In *The future of the presidency, journalism, and democracy: after Trump* (pp. 107-126). Routledge.

Brown, D. & Younes, R. (2023, 20 Dec). Meta's broken promises: Systemic censorship of Palestine content on Instagram and Facebook. *Human Rights Watch.* https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and

Bruns, A., Harrington, S., & Hurcombe, E. (2021). *Coronavirus conspiracy theories: Tracing misinformation trajectories from the fringes to the mainstream. Communicating COVID-19: Interdisciplinary perspectives,* 229-249.

Buntain, C., Innes, M., Mitts, T., & Shapiro, J. (2023). Cross-platform reactions to the post-january 6 deplatforming. *Journal of Quantitative Description: Digital Media, 3*

Chandrasekharan, E., Pavalanathan, U., Srinivasan, A., Glynn, A., Eisenstein, J., & Gilbert, E. (2017). You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech. *Proceedings of the ACM on human-computer interaction, 1*(CSCW), 1-22

Children First Act 2015, Part 1. (2015). *Dail Eireann.*

Coaston, J. (2018, 6 Aug). Youtube, Facebook, and Apple's ban on Alex Jones, explained. *Vox.* https://www.vox.com/2018/8/6/17655658/alex-jones-facebook-youtube-conspiracy-theories

Cohen-Almagor, R. (2016). Facebook and Holocaust Denial. *Justice,* 57, 10-16.

Corredor, E. S. (2019). Unpacking "gender ideology" and the global right's antigender countermovement. *Signs: Journal of Women in Culture and Society, 44*(3), 613-638.

Daniels, J. (2018). *Cyber racism: White supremacy online and the new attack on civil rights.* Rowman & Littlefield Publishers.

Darmstadt, A., Prinz, M., & Saal, O. (2019). The murder of Keira: Misinformation and hate speech as far-right online strategies. In M. Fielitz & N Thurston (Eds), *Post-digital cultures of the far-right* (pp. 155-167). Transcript.

Dawson, J. (2021). *This book is gay.* Sourcebooks, Inc.

Dawson, J. (2014, 4 Sep). Why my book is gay: And I'm proud of it. *The Guardian.* https://www.theguardian.com/childrens-books-site/2014/sep/04/this-book-is-gay-james-dawson

Díaz, Á. & Hecht-Felella, L. (2021). Double standards in social media content moderation. *Brennan Center for Justice at New York University School of Law.* https://www.brennancenter.org/our-work/research-reports/double-standards-socialmedia-content-moderation

Donovan, J., Lewis, B., & Friedberg, B. (2019). Parallel ports: Sociotechnical change from the Alt-Right to Alt-Tech. In M. Fielitz & N. Thurston (Eds), *Post-digital cultures of the far right: Online actions and offline consequences in Europe and the US* (pp. 49–65). Transcript.

Fanucci, B. (2023, 22 Mar). 29% increase in hate4 crimes and hate-related incidents in Ireland, Gardaí report. *GCN.* https://gcn.ie/hate-crimes-2022-ireland-gardai-report

Fitzgerald, C. (2023, 8 Aug). Explainer: Why is the far-right targeting Ireland's libraries? *The Journal.* https://www.thejournal.ie/library-protests-ireland-6135746-Aug2023/

Giansiracusa, N. (2121, 15 Oct). Facebook uses deceptive math to hide its hate speech problem. *Wired.* https://www.wired.com/story/facebooks-deceptive-math-when-it-comes-to-hate-speech/

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media.* Yale University Press.

Godzisz, P. & Viggiani, G. (2019). *Awareness of anti-LGBT hate crime in Europe*. Lambda Warsaw.

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), 1-15.

Guhl, J. & Davey, J. (2020). Hosting the 'holohoax': A snapshot of holocaust denial across social media. *The Institute for Strategic Dialogue*.

Innes, H. & Innes, M. (2023). De-platforming disinformation: Conspiracy theories and their control. *Information, Communication & Society*, *26*(6), 1262-1280. https://doi.org/10.1080/1369118X.2021.1994631

Heasman, A. (2023, 25   Jun). N/A. *Facebook*. https://www.facebook.com/100088622358284/videos/172480498979574

The Santa Clara Principles: On transparency and accountability in Content Moderation. https://santaclaraprinciples.org/

Jhaver, S., Boylston, C., Yang, D., & Bruckman, A. (2021). Evaluating the effectiveness of deplatforming as a moderation strategy on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1-30.

Kirk, R. & Schill, D. (2024). Sophisticated hate stratagems: Unpacking the era of distrust. *American Behavioral Scientist*, *68*(1), 3-25.

Klein, E. (2018 July 20). The controversy over Mark Zuckerberg's comments on Holocaust denial, explained. *Vox*. https://www.vox.com/explainers/2018/7/20/17590694/mark-zuckerberg-facebook-holocaust-denial-recode

Leidig, E. (2023). *The women of the far right: Social media influencers and online radicalization*. Columbia University Press.

Linehan, A. (10 May 2023). Performers disappointed after carlow drag show cancelled following online complaints. *GCN*. https://gcn.ie/carlow-drag-show-cancelled/

Mamié, R., Horta Ribeiro, M., & West, R. (2021). Are anti-feminist communities gateways to the far right? Evidence from Reddit and YouTube. *Proceedings of the 13th ACM Web Science Conference 2021*, 139-147.

Martiny, C. & Lawrence, S. (2023). A year of hate: Anti-drag mobilization efforts targeting LGBTQ+ people in the US. *ISD.*

McGreevy. R. (2023, 16 Sep). Rival protests over trans rights take place in Dublin city. *Irish Times.* https://www.irishtimes.com/ireland/social-affairs/2023/09/16/rival-protests-over-trans-rights-take-place-in-dublin-city

META. *Hate speech: Policy details.* https://transparency.fb.com/policies/community-standards/hatespeech/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2Fhate_speech

Mudde, C. (2019). *The far right today.* John Wiley & Sons.

Mulhall, J. (2019, 4 Oct). Deplatforming works: Let's get on with it. *Hope Not Hate.* https://hopenothate.org.uk/2019/10/04/deplatforming-works-lets-get-on-with-it/

O'Connor, C. (2023, 12 Sep). First libraries, now schools are under attack for teaching about LGBTQ+ identities. *Irish Examiner.* https://www.irishexaminer.com/opinion/commentanalysis/arid-41224576.html

Peck, H. (2023, 20 Dec). Exposing social media platforms' failures to protect their users. *Global Witness.* https://www.globalwitness.org/en/blog/exposing-social-media-platforms-failures-to-protect-their-users/

Rajan, B. & Venkatraman, S. (2021). Insta-hate: An exploration of Islamophobia and right-wing nationalism on Instagram amidst the COVID-19 pandemic in India. *Journal of Arab & Muslim Media Research*, *14*(1), 71-91.

Rauchfleisch, A. & Kaiser, J. (2021). Deplatforming the far-right: An analysis of YouTube and BitChute. *Available at SSRN 3867818*, 1-28.

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, *35*(3), 213-229.

Robertson, A. (2022, 5 Aug). Alex Jones must pay $4.1 million for spreading Sandy Hook Conspiracy Theory. *The Verge.* https://www.theverge.com/2022/8/4/23292425/alex-jones-sandy-hook-defamation-trial-verdict-fine

Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, *35*(3), 213-229.

Romano, A. (2016, 20 Jul). Milo Yiannopoulos's Twitter ban, explained. *Vox*. https://www.vox.com/2016/7/20/12226070/milo-yiannopoulus-twitter-ban-explained

Saleem, H. M. & Ruths, D. (2018). The aftermath of disbanding an online hateful community. *arXiv preprint arXiv:1804.07354*.

Saresma, T., Karkulehto, S., & Varis, P. (2020). Gendered violence online: Hate speech as an intersection of misogyny and racism. In *Violence, gender and affect: Interpersonal, institutional and ideological practices* (pp. 221-243). Springer.

Sellnow, T. L., Parrish, A., & Semenas, L. (2019). From hoax as crisis to crisis as hoax: Fake news and information disorder as disruptions to the discourse of renewal. *Journal of International Crisis and Risk Communication Research*, *2*(1), 121-142.

Slothouber, V. (2020). (De)trans visibility: Moral panic in mainstream media reports on de/retransition. *European Journal of English Studies*, *24*(1), 89-99.

Sontag, S. (1989). *AIDS and its metaphors.* New York.

Urman, A. & Katz, S. (2022). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society*, *25*(7), 904-923.

Van Dijck, J., de Winkel, T., & Schäfer, M. T. (2023). Deplatformization and the governance of the platform ecosystem. *New Media & Society*, *25*(12), 3438-3454.

FREEDOM OF EXPRESSION, DISINFORMATION AND HATE SPEECH: THE CASE OF COLOMBIA'S VICE PRESIDENT, FRANCIA MÁRQUEZ

Lina Uribe-Henao
/ Universidad Icesi, Colombia

Maria Juliana Soto-Narváez
/ Universidad Icesi, Colombia

Mauricio Guerrero-Caicedo
/ Universidad Icesi, Colombia

Presentation of the case: Francia Márquez and disinformation

In 2022, for the first time in Colombia, an Afro-descendant woman became Vice-President of the Republic. Her name is Francia Márquez Mina. Before she was nominated as vice-presidential candidate of the coalition that came to power (the *Pacto Histórico*), Márquez participated in a party consultation as a pre--candidate. Although she came in second place, she obtained remarkable electoral support and achieved a vote count that surpassed that of candidates with a more established political background.

Before her involvement in the executive branch, Márquez had emerged as a social leader committed to defending environmental rights in her native territory: Suárez, Cauca, a vulnerable region in southwestern Colombia. Her struggle led her to adverse situations, such as being forcibly displaced from her territory and being declared a military target by paramilitary groups eight years before she became vice president.

Márquez's arrival on the national political scene led to the proliferation of disinformation narratives that, since 2021, began to disseminate distorted representations of her identity and trajectory as a social leader, Afro-descendant woman, and leftist political figure. These narratives, based on prejudices of class, race, gender, and political ideology, contributed to creating and strengthening hate speech against her.

Smear strategies through disinformation began to be orchestrated and materialized by people deliberately dedicated to it. Social networks, in particular, were the main channel for the massive dissemination of misleading content, such as manipulated images, false videos, and other viral publications that contributed to feeding an atmosphere of distrust and hostility towards the candidate.

At the same time, media outlets dedicated to checking information tried to verify the dubious publications they saw online or that arrived at their channels with the express request to check them. *ColombiaCheck*, the first portal in Colombia dedicated mainly to fact-checking, published between December 2021 and December 2023, a total of 19 contents verifying disinformation in which Francia Márquez was the protagonist. If those linking her are included, the number rises to 94.

On the other hand, the Detector of Lies, a section of the portal *La Silla Vacía* dedicated to the verification of data, published in that same period 20 fact-checks of disinformation with Francia Márquez as the protagonist. Thus, between the two media, 39 contents verified disinformation in which Márquez was directly linked.

The digital era has posed new challenges for professionals dedicated to the production of information (such as journalists and academics), as well as for society in general, which faces the omnipresence of online disinformation. This phenomenon contributes to the polarization of discourses, the radicalization of opinions, the misinformation of the population, and, indirectly, to an erosion of trust in the journalistic profession and the work of the media (Rodríguez Pérez, 2019).

One of the greatest difficulties in addressing misinformation lies in the tendency of users to select sources of information that confirm their convictions (Lotero-Echeverry, Romero-Rodríguez, & Pérez-Rodríguez, 2018). In Colombia, one of the best-known cases of its disinformation strategy was the 2016 plebiscite, in which the citizenry voted overwhelmingly against the approval of the peace agreements between the government and the FARC guerrilla, motivated, in part, by the dissemination of misrepresented and erroneous content about the proposals and the scope of the agreements.

According to Rodríguez Pérez (2019), one of the reasons why it is necessary to talk about disinformation is that this phenomenon is not reduced to the publication of fraudulent informative content (or fake news), but encompasses both this concept and misleading content, hate speech, deliberate false speeches or unintentional informative errors by media or journalists. "Disinformation, therefore, is the distortion of information through the dissemination of falsified news that misleads the final recipient" (Rodríguez Pérez, 2019: 3).

Additionally, from an ideological perspective, disinformation seeks to favor certain positions or candidates (Allcott & Gentzkow, 2017, cited by Rodríguez Pérez, 2019), while simultaneously weakening public trust in the media. Added to this are aspects related to intersectionality: "hate speech can be motivated by reasons of race, color, ancestry, national or ethnic origin, ideology, age, disability, religion, sex, gender, gender identity, sexual orientation, and other personal characteristics or conditions" (Amores et al., 2021: 104), elements of the narrative intended to be viralized against the targeted individual.

In the case of Francia Márquez, the disinformation verified by the check portals that were taken into account for this chapter occurred in three key moments of her political career: her candidacy as presidential pre-candidate (3 checks), her vice-presidential candidacy (18 checks) and her first months in office (18 checks).

Social networks, especially Twitter and Facebook, emerge as the main vectors for the dissemination of disinformation, followed by platforms such as WhatsApp, Instagram, TikTok, and Kwai. The most recurrent disinformation narratives focus on representations that present Márquez Mina as a member of guerrilla groups, a fraudulent contractor for the State, and a practitioner of witchcraft and Santeria (Saavedra, 2022).

Of the 39 checks made by both media in the aforementioned period, the majority (28) were labeled as false, while five were "misleading", three were "questionable", one was "true" and another was "true, but with caveats". The remaining check was labeled as "multiple checks", titled "This is how the misinformation about alleged contracts of Francia Márquez with the State has moved".[1]

Some themes of the checkups and their narratives

"Here you have Francia Marquez, she is not as harmless as she seems".[2] This text, which accompanies the image that circulated during the 2022 presidential campaign and was checked by *ColombiaCheck* and *La Silla Vacía*, can give some initial clues about the axes of the narrative that strongly oriented the smear strategy that was deployed against Francia Márquez, at that time candidate. This type of contents that address imaginaries about politics tends to go viral and is a strong point in the disinformative agendas (Hernández-Rodríguez & Londoño, 2023).

1. Translation from authors of "Así se ha movido la desinformación de supuestos contratos de Francia Márquez con el Estado".
2. Translation from authors of "Aquí tienen a Francia Márquez, no es tan inofensiva como parece".

Freedom of expression, disinformation and hate speech:
The case of Colombia's vice president, Francia Márquez

Figure 1. Check an image presenting Francia Márquez as a member of the ELN guerrilla group

Note: This check, from the *ColombiaCheck* portal, was published on March 13, 2022, after an image started circulating on Twitter, and was rated "False".

Figure 2. Lie Detector checks the image presenting Francia Márquez as a member of the ELN guerrilla group


Note: This check, from the portal *La Silla Vacía*, was also published on March 13, 2022, and was rated as "False".

Electoral processes understood as confrontation install logics of dispute that can overflow ethical and legal frameworks, and disinformation is configured as a powerful and historically recognized tool to distort the rival and mark polarization scenarios (Acevedo, Quiroz & Villabona, 2022). In this sense, Francia Márquez was seen by her adversaries as an opponent who gathered sensitivities that could generate massive support, but also as a figure that could represent, through false information, old hatreds, fears, and biases. Therefore, pointing out that she belonged to the ELN guerrilla activated a narrative against her and displaced her from her political space, placing her in one of tension within the public's imagination.

In this same line, there are other images and texts checked by the two media, in which Francia Márquez is related to the FARC guerrilla. Once again, the narrative described above is activated, which seeks to reinforce a bias of disapproval and fear present in a part of the citizenry concerning the guerrilla groups. This issue had already been staged with particular nuances during the 2016 plebiscite, mentioned at the beginning of this chapter, which was also marked by widely known and referenced disinformation strategies.

On the other hand, the image of a tweet with false information that was checked and analyzed allows following other routes of the narrative against candidate Márquez. The text brings up a topic that has served as a ghost and distractor in several electoral processes in Latin America: The case of Venezuela and the possibility of repeating its experience. This text also synthesizes several ideas that allude to imaginaries already fixed in digital audiences in Colombia: The criticism and destruction of the capitalist system by the leftist ideology and the fear of living like the perception of Venezuela.

Figure 3. Checks to the disinformation linking Francia Márquez to the Farc-ep guerrilla group
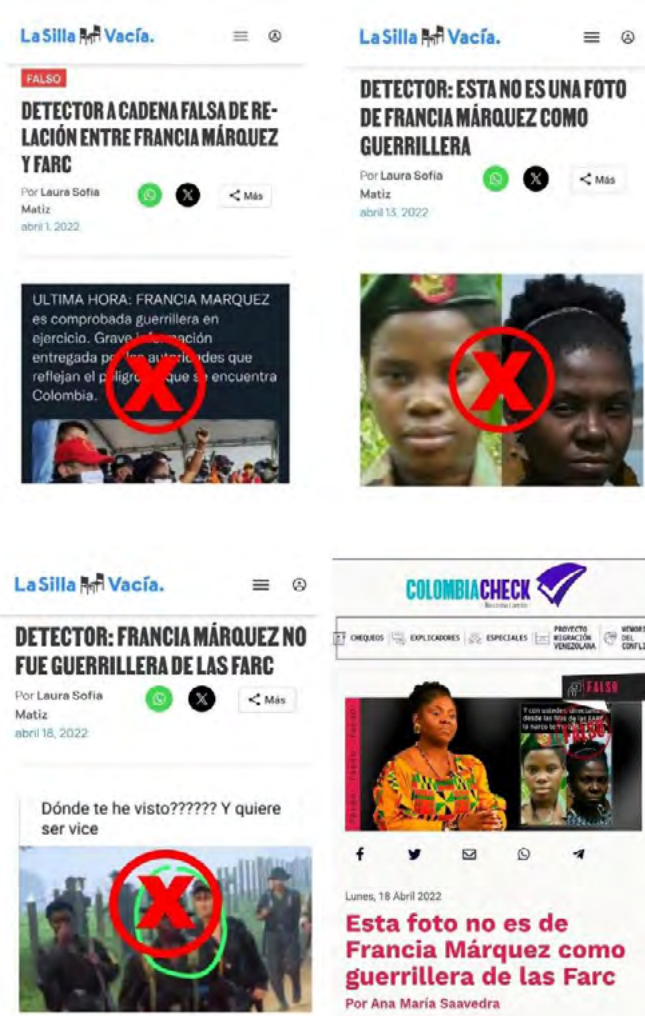
Note: Checks published in the portal *La Silla Vacía* (April 1, 13, and 18, 2022) and *ColombiaCheck* (April 18, 2022).

Regarding the composition of the image, it is observed how this type of montage simulates well the tone of writing used in X or Twitter, managing to generate a strong impact on the communities of like-minded users who receive them, given its simplicity and its ability to connect two theses in a logic of cause and effect. Another aspect that is important to point out in this case, and which is referenced by *ColombiaCheck*'s check, is that Álvaro Uribe Vélez, one of the former presidents who have generated the most polarization in Colombia, retweeted the image and then deleted it.

This action may have had a stronger effect in terms of message efficacy, as Álvaro Uribe Vélez, as a figure of "authority on the network" (Calvo and Aruguete, 2020), is a credibility reference, coupled with the fact that his political affiliation is on the opposite side of Márquez's ideological spectrum.

Figure 4. Check a tweet that held Francia Márquez responsible for an idea that incited the destruction of the capitalist system

Note: The check was published on March 25, 2022, by the portal *La Silla Vacía*. That same day, former president Álvaro Uribe Vélez retweeted it with the message "This is what Doña Francia will tell my fellow countrymen in Antioquia!". The fake image was dated January 5, 2019.

The review of checks also highlights a set that verifies information about the financing of Francia Márquez's campaign, her alleged relationship with the State as a beneficiary of aid and operator of contracts, and her lifestyle. Once again, the objective of distorting already established ideas and prejudices emerges, which can strengthen a discourse of hatred and rejection with a new narrative that links them. These prejudices are related to topics ranging from corruption to pointing out supposed ideological contradictions between what Marquez represents publicly and his personal actions.

Specific examples of the aforementioned can be seen in the checks on alleged properties acquired by Márquez in exclusive sectors of Colombia and the United States when she was already serving as vice president.

Regarding the racial, two checks disprove videos of a tiktoker who, pretending to be a daughter of Francia Márquez, shares in a satirical tone her opinion regarding a position of victimization on the part of her supposed mother. This content went viral and was replicated by users who assumed that the protagonist was Marquez's daughter. Several complex aspects are at play in terms of discrimination, as the discourse is generated from the intersection of her gender (limited to the role of mother), race (she is blamed for playing the victim of the racism suffered), and class (by criticizing the reiterations of her humble origin).

Figure 5. Checks on misinformation about the alleged lifestyle of vice president Francia Márquez
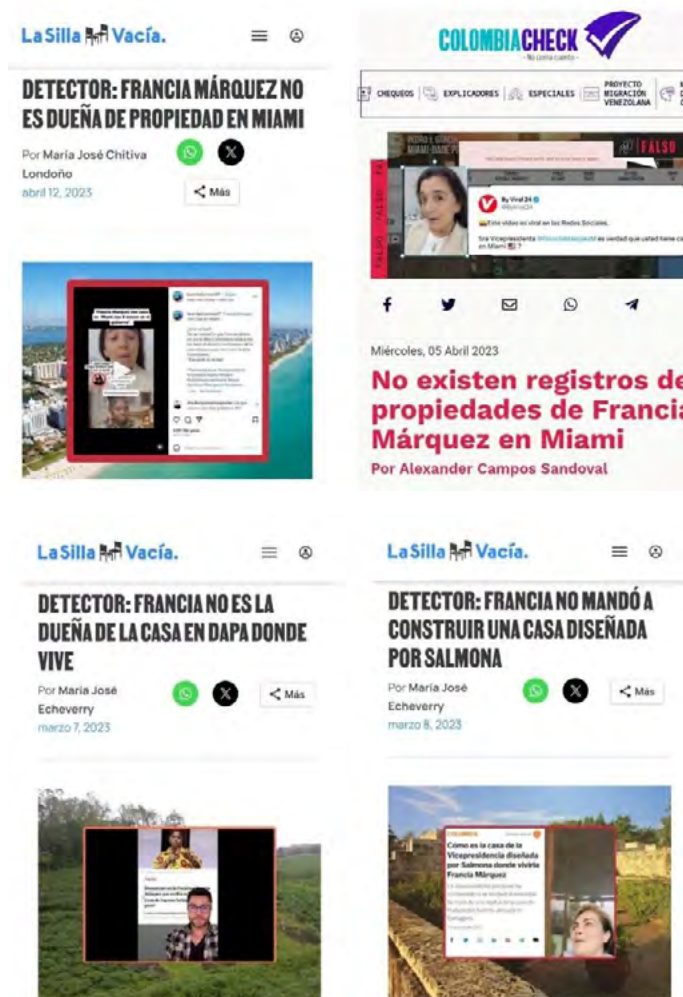
Note: Checks published in the portal *La Silla Vacía* (April 12 and March 7 and 8, 2023) and *ColombiaCheck* (April 5, 2023).

Figure 6. Checks on lying videos in which Francia Márquez's supposed daughter spoke

Note: Checks were published in the *ColombiaCheck* portal and *La Silla Vacía* on May 25, 2022.

Finally, there is evidence of a narrative of disinformation linked to religious, moral, and ethical aspects, which are addressed with decontextualized text images and false information. In this case, the topics chosen are abortion and Santeria, treated from a perspective that poses dilemmas, judgments, and debates in a light and biased manner. In this type of content, a strategy based on the crossing of prejudices to find a particular meaning in the political struggle is pointed out once again.

At a narrative level, we find in this type of content an orientation towards the creation of representations that can begin to function as a matrix for the management of more content along these lines. Installing the idea that Marquez does witchcraft opens a very wide space of meanings in a society like the Colombian one.

Figure 7. Check a publication that went viral on Facebook and Twitter, which suggests that Francia Márquez practices witchcraft and Santeria

Note: Check was published in the *ColombiaCheck* portal on June 16, 2022.

The blurred boundaries of freedom of expression

The growing concern about the impact of disinformation narratives, both at the individual level and for democratic systems, has prompted various sectors, ranging from academia and the press to human rights organizations, the UN, and governments, to undertake pedagogical actions, as well as research and regulation that have put at the center the question of what are the limits of freedom of expression in the dynamics of interaction that occur in digital environments.

In this regard, it is pertinent to recall that freedom of expression is a fundamental right protected by Article 19 of the *Universal Declaration of Human Rights*.[3]

According to the manual *Hate Speech*, published by the international NGO Article 19, "International human rights law requires States to guarantee to all persons the freedom to seek, receive or impart information and ideas of any kind, regardless of frontiers, through any medium of the individual's choice" (2015: 6).

The right to freedom of expression is further enshrined in Article 19 of the International *Covenant on Civil and Political Rights* (ICCPR).[4] In that instrument it is described as follows:

> 1. No one may be subjected to interference on account of his opinions.
>
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive, and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or print, in the form of art, or through any other media of his choice.
>
> 3. The exercise of the right provided for in paragraph 2 of this article entails special duties and responsibilities. It may therefore be subject to certain restrictions, which must, however, be expressly fixed by law and necessary to:
>
> a. To ensure respect for the rights or reputation of others;
>
> b. The protection of national security, public order, public health, or morals.

In particular, the ICCPR defines in its Article 20 what could be considered a limit to freedom of expression, establishing it as a non-absolute right, meaning that it can be restricted by States through legal mechanisms.

---

3. See the Universal Declaration of Human Rights: https://www.un.org/es/about-us/universal-declaration-of-human-rights
4. See the International Covenant on Civil and Political Rights: https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

Article 20

> 1. Any propaganda for war shall be prohibited by law.

> 2. Any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence shall be prohibited by law.

In the case of Colombia, the 1991 Constitution enshrines the right to freedom of expression in its Article 20:

> Every person is guaranteed the freedom to express and disseminate his or her thoughts and opinions, to inform and receive truthful and impartial information, and to find mass media. These are free and have social responsibility. The right to rectification under fair conditions is guaranteed. There shall be no censorship.

Another right that is directly related to the question of the limits of freedom of expression in digital environments is the right to equality and non-discrimination for all persons, enshrined in Article 1 of international human rights law.[5] Thus, discrimination is understood as:

> 1. Any distinction, exclusion, restriction, or preference against a person.

> 2. That is based on a protected characteristic, recognized under international human rights law.

> 3. And which has the purpose or effect of nullifying or impairing the recognition, equal enjoyment, and exercise of human rights and fundamental freedoms in political, economic, social, cultural, or any other field. (Human Rights Committee, 1989, cited by Article 19, 2015).

Therefore, the right to freedom of expression and the right to equality is at stake in the face of the phenomenon of hate speech on digital platforms (Igareda, 2022), whose manifestations occur, many times, under the figure of anonymity or, as mentioned above, through deliberate or orchestrated

---

5. See the Universal Declaration of Human Rights: https://www.un.org/es/about-us/universal-declaration-of-human-rights

Freedom of expression, disinformation and hate speech:
The case of Colombia's vice president, Francia Márquez

narrative strategies with clear purposes of discrediting, discrimination or dissemination of lies.

Responses to hate speech are what are known as limits or restrictions on freedom of expression. As stated in Article 19's "Hate Speech" Handbook, "Hate speech is an emotive concept, so there is no universal definition accepted by international human rights law. Many people will claim that they can identify what is 'hate speech' when they perceive it. However, the criteria for identifying it are often elusive or contradictory" (2015: 9).

The danger, however, is that when seeking to protect people from this type of speech, freedom of expression may be overly or abusively limited. Although there is no consensus on what to label under the concept of "hate speech", it has made it possible to encourage the search for spaces free of the violence that promote equality and dignity. Likewise, it has shown how discriminatory stereotypes of gender, ethnicity, nationality, sexual orientation, social class, and migratory status, among others, continue to be perpetuated through language, as an action that creates subjectivities and realities.

Based on the guidance formulated by the then UN Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Frank La Rue, in his annual report to the General Assembly, A/76/357, September 7, 2012, Article 19's "Hate Speech Manual" proposes a typology of "hate speech" according to the level of seriousness it would entail:

> 1. "Hate speech" that should be banned.
>
> 2. The "hate speech" that can be banned.
>
> 3. The legal "hate speech" should be protected under Article 19(2) of the ICCPR.

Although an analysis of the full typology is beyond the scope of this chapter, it is pertinent to at least mention the existence of such exercises which, from a civil society perspective, highlight the complexities of a concept such as "hate speech".

Although the 39 expressions checked by *ColombiaCheck* and *La Silla Vacía* with information related to vice president Francia Márquez do not reach the highest level of incitement to genocide, they do move in a gray zone between speech that "can be restricted" and speech that "must be protected" by freedom of expression, according with Article 19 typology (2015: 19).

Therefore, the set of checks analyzed is a sample of the contribution that this type of expression makes to the promotion of stereotypes and the stigmatization of certain groups or individuals, which causes a high level of concern, as they are undoubtedly problematic and require educational, political, social, and interdisciplinary measures to address their effects. In other words, freedom of expression has worrying nuances when it gives way to hate speech, which spreads due to the uncontrolled power of content on social media, leading some individuals to attack those they consider inferior (Di Fátima, 2023).

Disinformation and hate speech: A dangerous duo

In Colombia, a group of researchers in the field of behavioral sciences became interested in understanding the relationship between disinformation and psychology. According to information on their website SomosDIP (2024), in the "ConoceDIP" section, they set out to "understand the Colombian population and the particular traits that can make them stronger in the face of the disinformation attacks that are inevitable in this new digital era".

In addition to an exhaustive literature review, they conducted a diagnosis of more than a thousand Colombians in different regions of the country through an instrument composed of two tests and two surveys.

In their work, the researchers of DIP (Detox Information Project) established four characteristics that are strongly related to Colombians' vulnerability to disinformation: "distrust of the other, taking away or subtracting human characteristics from the other, feeling discriminated against and being less tolerant to ambiguity (called high need for cognitive closure), that is, preferring to see everything in black and white" (SomosDIP, 2024).

As mentioned at the beginning of the chapter, among the most recurrent disinformation narratives against the figure of Francia Márquez are those that associate her with photographs of women in guerrilla groups and with being a practitioner of witchcraft and Santeria (Saavedra, 2022). If these narratives are read under the perspective of DIP, they appeal to at least two psychological characteristics that are related to the possibility that a person is more susceptible to "falling" into disinformation: dehumanization, defined as "considering other people or other groups of people as less human in emotional and/or cognitive terms" (SomosDIP, 2024) and distrust, a characteristic that prevents a person from establishing agreements with those he/she considers different.

As already mentioned, the figure of the guerrilla awakens a bias of disapproval and fear in Colombia, which is related to the formulation of the main enemy of Colombian society: the guerrillas. A narrative fed, in the last two decades, by the exacerbated appearance of information related to the actions of the guerrillas and the invisibility of other actors also victimizers, such as the paramilitaries, and, of course, the invisibility of their victims (López de la Roche, 2014). The role of the media helped to shape a process of dehumanization of the guerrillas, stripping them of their human characteristics, "recognizing them as amorphous beings, without identity, gender, feelings, or the possibility of second chances in society" (Sánchez-Prieto et al., 2022: 33).

This is why the false comparison of the image of vice president Márquez with any other guerrilla is such an easily spread strategy, since it awakens imaginaries that, probably, remain rooted in a large part of Colombian society. On the other hand, the narrative that associates Márquez with religious or mystical practices such as Santeria or witchcraft can be understood under the psychological characteristic of distrust.

Figure 8. Montage with a photo of Francia Márquez and Elda Neyis Mosquera, alias "Karina"

Note: Taken from the check titled Detector: This is not a photo of Francia Márquez as a guerrilla, from *La Silla Vacía* (Matiz, 2022).

Figure 9. A Facebook publication that misinforms about the event in which Francia Márquez was with Emberá indigenous people

Note: Taken from the check titled This photo does not prove that Francia Márquez is practicing witchcraft or Santeria, from *Colombiacheck* (Campos Sandoval, 2022).

In this case, distrust is linked to religious prejudices and the conservative values of Colombia as a Catholic country. Any other spiritual practice, such as the case of the mandala-shaped offering of the Emberá indigenous women that can be seen in the photo, or any symbol of syncretism that attracted attention in the clothing, speech, or gestures of the then vice presidential candidate from Colombia, Francia Márquez, could have reinforced distrust towards her by not seeing herself as the traditional Catholic woman whose most representative image is the Virgin Mary, a white, young and obedient woman.

Conclusions

The literature review and narrative analysis of the messages or expressions checked by *ColombiaCheck* and *La Silla Vacía* developed in this chapter reveal the relevance of incorporating different perspectives to study and counteract the problem of hate speech when it is interwoven with disinformation strategies, crossed by racial and class biases, and when they confirm stereotypes rooted in society, in this case, Colombian society. It is also pertinent to recognize formats, hoaxes, and feelings as variables to unravel the objectives of these discourses and their strategies (Gutiérrez-Coba & Rodríguez Pérez, 2023).

In this sense, it is necessary to address the problem of disinformation from a rights perspective, which seeks in a balanced manner the balance between the protection of the right to freedom of expression and the right to equality and non-discrimination.

Digital environments, where anonymity, the size of the audience, and the authority or lack of authority of a voice affect how hostile or offensive speech is moderated, eliminated, or allowed to circulate, require interdisciplinary analyses, at least among the communication, law, psychology, sociology, and computer science.

An example of this is the proposal that arises from behavioral sciences, as briefly mentioned in the chapter. From a psychological point of view,

behavioral sciences have expanded the framework of references and offered new perspectives for the design of interventions and strategies that seek to dismantle the chains of misinformation that lead to greater polarization and distrust among citizens.

Finally, the task of the portals in charge of verifying content that may be spreading hate speech camouflaged in misinformation requires an intersectional view, which does not focus on the misinformation itself, but rather unravels the content until finding the narratives that characterize it. and that incites hatred. In the case of Francia Márquez, as mentioned, these narratives were created around gender, social class, race, and political orientation biases.

References

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info*, (49), 98-124. https://dx.doi.org/10.7764/cdi.49.27817

Artículo 19. (2015). *Discurso de odio* [Manual]. https://bit.ly/4cYjtuM

Calvo, E. & Aruguete, N. (2020). *Fake news, trolls y otros encantos. Cómo funcionan (para bien y para mal) las redes sociales*. Siglo XXI.

Campos Sandoval, A. (16 de junio de 2022). Esta foto no demuestra que Francia Márquez esté practicando brujería ni santería. *ColombiaCheck*. https://bit.ly/49wHs1f

Campos Sandoval, A. (5 de abril de 2023). No existen registros de propiedades de Francia Márquez en Miami. *ColombiaCheck*. https://bit.ly/446jIjq

Chitiva Londoño, M. J. (12 de abril de 2023). Detector: Francia Márqeuz no es dueña de propiedad en Miami. *La Silla Vacía*. https://bit.ly/4azdv1A

Di Fátima, B. (2023). *Hate speech on social media: A global approach*. LabCom Books & EdiPUCE.

Echeverri, G. L., Rodríguez, L. M. R., & Rodríguez, M. A. P. (2018). Fact-checking vs. Fake news: Periodismo de confirmación como componente de la competencia mediática contra la desinformación. *Index. Comunicación: Revista científica en el ámbito de la Comunicación Aplicada*, *8*(2), 295-316.

Echeverry, M. J. (7 de marzo de 2023). Detector: Francia no es la dueña de la casa en Dapa donde vive. *La Silla Vacía*. https://bit.ly/49BK3H1

Echeverry, M. J. (8 de marzo de 2023). Detector: Francia no mandó a construir una casa diseñada por Salmona. *La Silla Vacía*. https://bit.ly/3xvEfl0

Gutiérrez-Coba, L. & Rodríguez-Pérez, C. (2023). Estrategias de posverdad y desinformación en las elecciones presidenciales colombianas 2022. *Revista de Comunicación*, *22*(2), 225-242.

Hernández-Rodríguez, J. C. & Londoño Pardo, O. (2023). Los contenidos, los actores y los intereses detrás de las noticias falsas. Un análisis de los bulos verificados en España y Colombia. *Comunicación y Sociedad*, e8518. https://doi.org/10.32870/cys.v2023.8518

Igareda, N. (2022). El discurso de odio anti-género en las redes sociales como violencia contra las mujeres y como discurso de odio. *Derechos y Libertades*, *47*, 97–122. https://nebulosa.icesi.edu.co:2144/10.20318/dyl.2022.6875

López de la Roche, F. (2014). *Las ficciones del poder. Patriotismo, medios de comunicación y reorientación afectiva de los colombianos bajo Uribe Vélez (2002-2010)*. Debate – Universidad Nacional de Colombia.

Matiz, L. S. (1 de abril de 2022). Detector a cadena falsa de relación entre Francia Márquez y Farc. *La Silla Vacía*. https://bit.ly/3VXFHqu

Matiz, L. S. (13 de abril de 2022). Detector: Esta no es una foto de Francia como guerrillera. *La Silla Vacía*. https://bit.ly/4az02Hk

Matiz, L. S. (13 de marzo de 2022). Detector: Esta foto de Francia como guerrillera del ELN es falsa. *La Silla Vacía*. https://bit.ly/4cQwJSd

Matiz, L. S. (18 de abril de 2022). Detector: Francia no fue guerrillera de las Farc. *La Silla Vacía*. https://bit.ly/3Ul4zYk

Matiz, L. S. (25 de marzo de 2022). Detector: Francia no trinó "tenemos que acabar con el capitalismo". *La Silla Vacía*. https://bit.ly/3TQtZeM

Matiz, L. S. (25 de mayo de 2022). Detector: Una mujer se hace pasar por hija de Francia Márquez para criticarla. *La Silla Vacía*. https://bit.ly/3vRlS9M

Palma Lozano, M. (25 de mayo de 2022). La mujer de este video no es hija de Francia Márquez sino una 'tiktoker' que sube contenido satírico. *ColombiaCheck*. https://bit.ly/3VZsEox

Rodríguez Pérez, C. (2019). No diga fake news, di desinformación: Una revisión sobre el fenómeno de las noticias falsas y sus implicaciones. *Comunicación*, (40), 65-74.

Saavedra, A. M. (12 de julio de 2022). Guerrillera, contratista y bruja: narrativas de desinformación contra Francia Márquez en la campaña. *ColombiaCheck*. https://bit.ly/49wn400

Saavedra, A. M. (13 de marzo de 2022). La mujer de la foto con pañoleta del ELN no es Francia Márquez. *ColombiaCheck*. https://bit.ly/4cUiETH

Saavedra, A. M. (18 de abril de 2022). Esta foto no es de Francia Márquez como guerrillera de las Farc. *ColombiaCheck*. https://bit.ly/3xyXQ3U

Sánchez-Prieto, V., Ospina-Deaza, J. C., & Clavijo-Poveda, J. (2022). Biopoder y máquina antropológica. La deshumanización y humanización de los actores armados en Colombia como tecnologías de poder. *Revista Austral De Ciencias Sociales*, (42), 27-46. https://bit.ly/44fzubT

SomosDIP. (2024). *ConoceDIP*. https://www.somosdip.com/conocedip

Freedom of expression, disinformation and hate speech:
The case of Colombia's vice president, Francia Márquez

HATE SPEECH IN ARGENTINE SOCIETY: ORIGINS AND CONTEMPORARY EXPRESSIONS

Sergio G. Grabosky
/ University of Salta, Argentina

Theoretical foundations

In these post-pandemic times, the digital buzz has increased the social murmuring and complaints that are augmented by online hate messages. The number of voices that make information and disinformation go viral produces content that vindicates hate crimes in the last civil and military Argentine dictatorship; hate speech is memorable (Grabosky, 2022). This nostalgic insight into the darkest passages of Argentine history is updated when identifying new social agents as targets of hatred and antidemocratic violence. Feminism, LGBTIQ+ community, speakers of inclusive language, marginalized social groups, targets of inclusive public policies – *planeros* according to the hate speech – are added to *zurdos*, a derogatory term that makes reference to and dismisses many sectors of society, mixing and distorting political affiliations. Therefore, *peronistas*, supporters of social organizations, union representatives, militants are built again as the enemy who has to be destroyed. It is not coincidence that the *kirchnerismo* is loathed because it is the political expression of the expansion of rights for the unemployed or informal workers. The implementation of public policies is not tolerated, in particular those policies that embrace respect for diversity and gender identity and laws enacted in that regard (National Congress of Argentina.

2012, May 23. Bill N° 26.743 on Gender Identity in Argentina; National Congress of Argentina. 2006, October, 23. Bill N° 26.150. National Program of Comprehensive Sexual Education; National Congress of Argentina. 2015 December 30. Bill N° 27.234. *Educar en Igualdad*: Prevention and Elimination of Gender Violence; National Congress of Argentina. 2019, January 10. Bill N° 27.499, Mandatory Training in Gender for workers of the Three Branches of Government).

The incorporation of health protocols to guarantee reproductive rights is criticized with hatred, for example, the National Plan of Prevention of Unintended Pregnancy in Adolescents, ENIA (Ministry of Health) and any other policy of workplace integration of the LGBTIQ+ community (National Congress of Argentina. 2021, July 8. Bill N° 27.636 Access to Formal Employment for Crossdressers, Transsexuals and Transgenders).

At the beginning of 2022, a social buzz was increasing with the importance that mainstream media adds to these speeches that are created on social media, reverberating them like an exhaust pipe that pollutes and destroys the democratic ecosystem. Apart from the media, there are also voices and ways of doing politics of allied groups, powerful groups of the mainstream economy, which are the center of the opposition of a Peronist government as the one led by Alberto Fernández, who had to handle with the new debt between Argentina and the International Monetary Fund taken by the previous government in charge of the engineer Mauricio Macri, in the context of a new global crisis, the covid-19 pandemic.

That buzz – hate messages, intolerance, exacerbation of the insecurity feeling that mainstream media builds in the news – together with the unhappiness originated by the economic crisis have created a breeding ground which paved the way for the attempt of magnicide with a weapon against the vice-president Ph.D. *Cristina Fernández* on 1st September 2022. Weeks before, her supporters were camping in *Recoleta*, opposite Cristina's house, to publicly support her figure in the closing arguments by the Attorney who investigated the case of *Grupo Austral* and requested imprisonment

of her and many high officials as well as banning them from public office. Anonymous voices and public figures' voices that were attacking *peronistas* and *kirchneristas, negros planeros*[1] and corrupted officials permeated social media and TV every day and every hour, eventually they have come up with the phrase: "It's them or us" (Lopez Murphy [(@rlopezmurphy], 2022; Walfeind, 2022).

The fact that the bullet did not hit the face of Ph.D. *Cristina Fernández de Kirchner* was due to a technical failure. Media was covering the supporters' camping; therefore, the attempt of assassination was broadcasted. However, many voices continued with the hatred by denying the attempt and accusing the victim of having invented it.

In 2023, 40 years after the return to democracy in Argentina, the country is divided by a serious cleavage, encouraged by hate speeches and crimes. In a new context of social and institutional violence, the beginning of 2023 is marked by illegal State repression in two provinces of the North of Argentina and finishes with the victory of the far-right candidate known as *libertarios* (libertarians), *Javier Milei*, as the elected president of Argentina.

On 25th May, the wick was fueled in Salta (Perfil, 2023; Brizuela, 2023). But the gunpowder deflagrated in the province of Jujuy, where teachers asking for better salaries and indigenous communities fighting for land, water and protesting against the Provincial and Constitutional Reform proposed by the government, were violently suppressed. These events broke with the democratic agreement of *Nunca Más* that had been in force since the Report written by the National Commission on the Disappearance of Persons (1984). In 2023, there have been some events: new illegal arrests by unidentified

---

1. The term *planero* and its alternative *choriplanero* are uttered by people belonging to conservative and middle-class sectors to attack to the unemployed, popular groups who receive any type of Social Security program. Since the government of Menem, the economic crisis was being mitigated by the creation of programs and plans of this type; however, the concept of *planero* uttered in a violent political way attempts to despise and attributes to the subject some ideas of laziness, dependence on the State and lack of skills to obtain things in life. That is why it is linked to the meritocracy concept that neoliberal and capitalist sectors account for as they ponder *peronismo* as the creation of poor political sectors and attached them to the necessity of receiving a social security plan as a way of being dependent on a political party. *Choripán* is a food that is said to be the payment to those lazy people and militants for attending a political event; it is a popular food that is characteristic of football stadiums, meetings among militants and protests.

police officers that drove private vehicles to detain demonstrators (*Ámbito.* 2023, June 27); the imprisonment and judicialization of the lawyer Nallar, just because he provided legal advice to demonstrators and questioned the Reform of the Government of Jujuy (Mamaní, 2023); tortious behavior and violence towards demonstrators who were followed and judicialized after protests (Brizuela, 2023).

Due to this situation, from the North of Argentina, *Tercer Malón de la Paz* decided to walk more than 1,550 km to arrive to Buenos Aires and request the Supreme Court of Justice to issue a statement regarding the unconstitutionality of the Reform in the Province of Jujuy, and then, they camped outside Congress to protest against repression (Grabosky. 2023, July 31). The presence of the representatives of indigenous groups in Buenos Aires triggered new hate speeches that some journalists created in the TV like Channel 13 (Página/12, 2023).

It is impossible to disconnect those hate speeches and acts that nowadays are intermingled in a context where dictatorship voices and a new State violence are present, with a repressive position of political groups towards the Argentinian right wing. In the election year, politicians belonging to *Juntos por el Cambio* (JxC) intended to run for office. The candidate *Horacio Rodríguez Larreta*, incumbent governor of the city of Buenos Aires, and his running mate Gerardo Morales, incumbent governor of the province of Jujuy, competed against Patricia Bullrich in the Open Primary, Simultaneous and Mandatory (PASO) Elections. She has openly proposed a strong-arm police method and the judicialization of social protest. In the run-off, she supported the elected president, who based his campaign on a constant attack to *kirchnerismo,* the State and the caste, on social media and TV, promoting the elimination of the public system of education and health, the privatization of retirement benefits, of scientific-technical institutions, of *Aerolíneas Argentinas*, of *YPF* under the motto of privatize as much as it can be privatized. His hate speech towards the political enemy makes allusion to the crimes of the last military dictatorship and denies the disappearance of 30,000 people. Besides, it is materialized in the threatening

gesture of rattling a chainsaw. The elected vice-president Villaruel is a negationist lawyer; this exhibits the political practice of violent and hate speech imposed by Milei and Villaruel who were largely supported by the Argentine society in the polls.

Under these circumstances, the analysis of hate crime must be done taking into account its roots in the Argentine history since the conquest and which continues to happen with impunity. We talk about *chineo*, a practice that still happens today in some regions of Argentina and refers to gang rape or individual rape to indigenous women and children perpetrated by creole men that "hunt" them to sexually abuse them. Even though it is linked to old traditions of sexual initiation of children who are part of the indigenous community, the *ramiada* or *chineo* is a hate crime that exhibits the sexism and power abuse over indigenous women's body, considered it as a trophy or land to conquer due to the colonial oppression (Cebrelli, 2022). *Chineo* is upheld in public media and social media through hate speeches that treat indigenous women as objects and animalizes them, according to Di Fátima (2023: 12) as one of the characteristics of the hate speech seeks to dehumanize the enemy. Therefore, the sexist invitation that a public official sent to participate in the folklore festival of *Trichaco salteño* with indigenous women pictures clearly indicates sexual connotations linked to *chineo* (Corbalán, 2022).

Methodological framework

Departing from the complex term hate speech, it is necessary to explain speech, hatred and hate crimes so as to understand a contemporary phenomenon linked to the *digital buzz*. This term describes the current context in which the Internet and the viralization of content on social media and mainstream media produces a breeding ground to exacerbate the production, spread and prosumerism of hate speech. The social buzz and voices on social media are materialized and strengthened as speech. Speech acts that harm the Other, an abominable, denied and excluded individual or collective subject.

Hate speech can be studied from the social discourse (Angenot, 2010) which grants us with a methodological framework suitable for the analysis of a complex speech, which enables to consider every utterance as a complete and institutionalized practice with a transphrastic meaning. This implies that the immanent view of speech has been overcome to consider the production and reception. The conditions for production, spread and reception of speech in the digital era are retrieved by the viralization, hashtags and ways of participating through comments, sharing, recreating and reusing content. A transdisciplinary view is adopted to consider many dimensions of hate speech: historical, cultural, legal, psychological, emotional and the dimension of materializing online writing. Those dimensions are not homogeneous but are complementary approaches which also make the perspective complex by establishing theoretical incompatibility aspects that come to grips with the analyzed characteristics of hate speech.

The corpus was chosen manually and without the use of automated analysis systems and was done by determining the topics of the public agenda on social media and mainstream media that occurred in the political context between 2022 and 2023. In this way, topics about political hatred toward peronismo were defined as objects for description and analysis because they acquired importance after the attempt of magnicide against Cristina. It is about a political and racial hatred (Grimson, 2017) targeted at *marrones, criollos* and *indígenas*; and political supporters of *kirchnerismo, peronismo* and *planeros*.

Besides, another topic is the remembrance of crimes perpetrated by the last military dictatorship as hate expressions against human rights and public policies of protection of minorities. Ideological and gender hatred involve derogatory terms that refer to *peronistas, zurdos*, members of the LGBTIQ+ community and indigenous women and their fight (for the right to legal, safe and free abortion, or the reproductive right and comprehensive sexual education) and Latin-American foreigners.

Multidimensional analysis of hate speech

Legal dimension of hate speech and its relationship with hate crimes

Hate speech has a legal dimension as it is linked to hate crimes. Miró Llinares F. (2016) argues that the term hate speech refers to the spread of utterances and forms of communication targeted at specific groups characterized by their race, religion, sexual orientation, disability, ethnicity, nationality, age, gender, social group, political affiliation or any other personal, functional or social characteristic or just for having an ideology they don't adhere to. This viral spread on social media is based on prosumerism. The author states that:

> Preoccupation is not about sending violent and hate material to thousands of people which is produced by specific groups, but it is about thousands of people sending hatred and violence in many forms, through social media, who interact between themselves and create, in many cases, a great social annoyance. Social media seems to be full of offensive and violent words in a context that maybe because of the potential anonymity that the communication is associated to or by any other factors related to the perception of lack of harm in the practices, it seems to enhance those forms of communication and/or, at least, increase their visibility. (Miró Llinares, 2016: 96)

The case of direct threat and forms of communication that seek to create in the other person a desire to harm – to incite direct violence – or to encourage a group of people to harm someone violently and physically is encompassed in the first category. It is included the tweet "It's them or us" (Lopez Murphy [(@rlopezmurphy], 2022) written in a context of political exacerbation around *Cristina Fernández de Kirchner* and posted by a group of violent people some days after the attempt of assassination. The fact that the tweet went viral and the attempt of a hate crime are linked almost explicitly to the time and context. However, there was not any judicial consideration of the hate speech uttered by the politician in the legal case for the attack.

In this way, we can consider incitement and threating acts which clearly involve certain willingness to commit or that somebody else commits violent acts without explicitly naming the targets but only their general characteristics. Even though the famous tweet by José Luis Esper (@jlesper) *"Carcel o bala"* (Jail or Bullet) in every circulation context defines his recipients, the dichotomy works as a speech that incites the materialization of violent, real and physical acts with the word bullet. In the discursive aspect, the phrase "Jail or Bullet", the term jail is known whereas the idea of bullet is new as if it were a way of overcoming the social struggle of delinquency and of *planeros* that protest in highways or take over *Recoleta* to defend their leader.

The second category proposed by Miró Llinares (2016) includes all those discourse forms in which violence incitement is indirect by positively assessing the commitment of violent and physical crimes perpetrated in the past or in the future over a specific or unspecific person. Vindicating the military dictatorship and its crimes as well as positive comments on repressive actions appear to be daily hate speech on Argentine social and traditional media during the analyzed period (Grabosky, 2022). This is observed when disclosing emblems of the Dictatorship, like the green Falcon to kidnap leftists, or explicit comments about death flights and torture that update the enemy to kill. Another example is the hate speech celebrated on social media and in some traditional means of communication, praising the police action of killing criminals in real contexts of execution or from behind. The case of the police officer *Chocobar* has gained support on social media by politicians and common users' accounts (Inofre. 2021 May 29; Oriland. 2021, May 27).

Subjective dimension of hate speech

In this part, another dimension of hate speech is considered. A subjective complexity that underlies hatred as a human emotion, attitude or feeling. Far from being controllable, emotions are linked to the unconscious which is out of our control. In the first place, we resort to Kristeva (1985) who defines the abject in the construction of subjectivity around the object that is left behind, that fallen object that is rejected, beyond the limits of what

is tolerable. The abject implies borders, an ambiguity space, which makes a distinction between "I" and what threatens the self. The abjection leads us to the archaism of pre-objectal relationship, in the immemorial violence with which a body becomes separated from another body in order to be. It resembles that obscurity where the shape of the object is lost and where the unmeasurable feeling acts. The author gives another significance to the abjection regarding the culture. Departing from the concept of abjection itself, the self can be discovered, who is not anybody else but the abject. That subject that knows the value of inaugural lost that defines all the objects which, ultimately, end up finding in the literature – the culture – the signifier of the abjection itself. Kristeva argues: "Every abjection implies the recognition of the fundamental lack of the self, the sense, the language, the desire". (1985: 12)

Following that overview, Winograd and Natale (2020) analyze hatred as something that has been part of humanity since its origins. Besides, taking into account Freud and Lacan's psychoanalysis, they delve into how the psychic constitution of the subject is linked to aggressiveness. They point out that according to Freud, the choleric actions targeted to the exterior comply with the discharging function that relieves the living being because it maintains constant brain arousal. There is also a relationship between group aggression as regards the psychic apparatus and its relationship with the exterior. Aggression or destruction is displayed as psychic results of the death drive. Aggressive impulses are the less damaging path because the aggression towards the inner self is not healthy whereas aggression towards the outer self relieves the living being. From this point of view, human being's aggressive tendencies cannot be suppressed. Attempts to express feelings of brotherhood and identification would be ways of mitigating the constitutive aggressiveness of the subject.

From this subjective dimension, the relationship between political hate speech towards *peronismo* can be explained, which has been considered as the structure of political hatred in Argentina. *Cabecita negra* represents the limit that a racist society wants to leave out, exclude or that is intensified through political signs of the racialized term (Grimson, 2017).

The underlying racism in the term negro that in Argentina not only makes reference to indigenous people, creoles, poor, Peronists, black, but also turns into xenophobia towards Paraguayans, Venezuelans, Bolivians and immigrants coming from Latin America. These immigrants are conceived as threats, a danger that deserves hatred and who could not be accepted. This situation is different for foreigners inside the country as it had happened in the migrations at the beginning of *peronismo*. There are repeated topics on social media. Thousands of users consume, resend viral messages, comment and recreate. Feelings and attitudes, repression against struggles that arise from new contexts of exploitation and disparities have led, as it was already mentioned, to the perpetration of crimes like the attempt of assassination to Cristina Fernández de Kirchner. We could interpret this attempt as the almost direct consequence of what Özlem Alikılıç, Ebru Gökaliler, İnanç Alikılıç define as shitstorm (2023: 98), which always results in a hate crime.

Online writing dimension

In this dimension, we consider writing as a way hate speech expression and its link to hatred as an emotion. Online writing is vague, unstable, nomad, anonymous and viral. That is why it favors this type of radicalized and hating content turning from a buzz, a rumor and a shout to the posting and circulation of a new way of political expression, many of them being antidemocratic (Amores et.al., 2021), which fulfills a hatred that grows in a country where there is a constant increase of social inequality with huge problems of poverty and in a pessimist context where cyclic economic crisis the marginalized who have poor living conditions. Giorgi and Kiffer establish that:

> Hatred burns as the fire of afflictions that harm the discursive tissues, not only in its annihilating brutality that lurks in language's own inability to speak, but also in its vindicating force to remake the living regimes of inscription and and its forms and expressions of a power to say. (2020: 9)

The relationship between hatred as a feeling and as a speech finds in the word online a space of production and viralization that contributes to the novelty of hate speech: that viralization and immediacy in which messages containing hatred as a feeling that struggles to be expressed circulate.

Giorgi also affirms that:

> Contemporary hatred – its strong nature – its unequivocally political quality, its temporary nature and myriad memories – is basically a written hatred: a hatred that is written in new territories, especially online ones, and that mirrors a radical transformation of writing, to which it cannot be split. Hatred has two sides: a political side with a new attraction in the democratic ways of expression when uttering new statements and subjectivities; also, it has a radical reconfiguration of the written world, its technologies, its circuits, its speakers and audience. (2020: 20)

Hatred is posted and goes viral. On social media, new voices can be expressed through online circuits that produce a change, an irritation in the social world. The author analyzes the tension as a word that defines the social mood in a context of polarization regarding the two governments in charge of *Cristina Fernández de Kirchner* (*Argentina* 2008-2015). Tension is a word that refers to the body, the muscle contraction that implies that the semantic movement is felt as an emotion and in the body. At that stage, tension was placed in the government, against some representatives of the oligarchic sectors or mainstream media. But also, the tension attacked the figure of the president *Cristina*, who suffered: (...) a verbal violence that recovers racist, sexist and classist voices, and that was fed by a long-standing anti-Peronist tradition" (2020: 22). The author characterizes Mauricio Macri's government as a space of enablement of racial, classist and xenophobic rhetoric as part of government strategies that are opposed to segregationist imaginaries against the idea of democratic inclusion.

Another aspect of hatred as a human emotion is that it can be both a destructive or annihilating force against the enemy and a creative and moving

force towards a social change. Kiffer (2020) puts forwards the dimension of political attitudes because bodies interact with concepts, ideas, values, ways of living, arts and writing. Political attitudes modify body shapes, gestures as lines of those vibrating bodies, of those words that seek to enter the discourse and delimit some political attitudes.

If we think in a historical perspective:

> Political attitudes that arise from the vindication of minority groups untidy, disturb and request that we study a great part of repressed national myths and political concepts. In the case of Brazil, miscegenation, an equal society among men and women, whites and blacks, but that is the crucial and complex point, they also seek to revalue those things that as an effect of repression designs an apparent calm and/or peaceful atmosphere for some time and in a particular society. (2020: 87)

The concept of Deviation that Kiffer adopts from Glissant E. allows to understand the phenomenon of living together but separately. He continues analyzing: Attitudes that are outside the neoliberal system can rise with chaotic force as strong and thrilling liberation and release because they were never treated, welcomed or received in that being or "living together" but separately (2020: 88).

What is currently happening in the North of Argentina – indigenous people protesting against the reform of the Provincial Constitution which was put forward by the governor Gerardo Morales without advice and in which some specific articles go against the right of the communities over the land – is the result of living together but separately, in a province – in a country – where structural racism and social and racial inequality have existed for a long time.

As opposed to this destructive hatred, feminist struggle, which also produces hate speech to revolutionize sexist oppression, can be considered a creative force of that antipatriarchy speech that tends to create a new way of politeness and being together. Negative hate speech, which destroys the good does not create but disposes victims. Kiffer argues:

...how to place oneself in that territory of struggle and rising of sub-jectivities that imagine democracy as a constant segregation and war? No, certainly, with some texts without hatred, with peaceful writings and education of the civic passions. Maybe, a mistake of our times is imagining a democratic subject as a "free of hatred" subject who is able to overcome passions with consent and dialogue, where writing would comply a crucial role in that education of passions to build an abstract or ideal civility. I think, on the contrary, about writings that promote hatred as a political attitude to create new shared spaces, to produce other pictures of collectiveness, as a kind of counter-offensive to escape from the uses of hatred as a way of affirming previous identities and extensive restoration of a mythical order (2020: 66).

The example that the author analyzes is the phrase in the *Ni una menos* demonstration: *Al patriarcado lo hacemos concha* (we break the patriarchy).

...Concha as a reinvention and destruction at the same time, to hate the patriarchy and to hate the *Macho*, and to turn that hatred into an eman-cipatory line, a body mutation *("hacer concha")* and subjectivities as an exercise of democratic imaginaries. What is hated is the grammar of violence that is the patriarchy. (Kiffer, 2020: 67)

It is coincidence that in Argentina, hate speech attacks the feminist struggle and its supporters. Traditional press tends to show feminist hate speech, writing and paintings on churches and monuments to create tension in the protests of *Ni una menos* or *Encuentros de Mujeres* with some values and democratic respect that only hide the desire to mute the rebel's shouts. The dimension of dissident bodies that are on the streets has always been a rea-son for sexist and media reactions, too.

Online writing which is pierced by hatred as a destructive and creative force is what puts new subjects of language immersed in these speeches into ten-sion. This writing places the Other that must be wretched, destroyed in a culture where violence is used and is a means of sending anonymous, harmful and antidemocratic messages.

We have already affirmed that the Other, deeply rooted in the national spirit, is the indigenous, the worker, the unemployed person, the Peronist, the *zurdo*. But it is also any other minority, as it can noticed in the systematic attack against supporters of feminist and LGTBTIQ+ movement and vegan activists that attack (and attacked by) *gauchos* in the premises of *La Rural* (Domínguez. 2019, July 31) as a main slaughterhouse, which serves as a symbol of that national spirit that has loathed what is different and finds in veganism the Other to be destroyed.

Historical and cultural dimensions

The historical dimension rebuilds the origins of hatred in Argentina. Hate speech goes through our history, as it is not new and has not arisen from the Internet. All the contemporary cultural expressions of those speeches are faced by the problem of construction of the National State that was forced in the destruction of the barbarian, the struggle for the construction of a white, Eurocentric, Christian and patriarchal civilization.

Why are the historical and cultural dimension studied together? We acknowledge that the artistic representations, in particular, literature enables to rebuild the roots of hate speech. Di Fátima (2023) contends that:

> As a communicative act, the roots of hate speech are the codes and values of a particular culture (Matamoros-Fernández & Farkas, 2021). These are only some of the challenges. Empirical studies based on Big Data show that detecting hate speech on social media is difficult (Miranda et al., 2022). Indeed, haters mobilize numerous subterfuges to obscure their intentions. For example, haters can use irony, humor, and satire to disguise a violent narrative (Schwarzenegger & Wagner, 2018).

This is the historical perspective that will enable us to find the roots of those cultural codes and values that are shown in the literary representations that make up the national literary corpus. The literary discourse allowed and allows to circulate hate speech camouflaged by the use of irony, humor, satire, resources that turn it into acceptable and that even now can lie to the algorithm.

The literary series allows us to consider the past of *hating* speeches. Piglia (1993) invents the metaphor: Argentina in pieces. The polysemy accounts for the destruction, turn-overs, violence and hatred along history. Hatred to the indigenous, to the immigrants, to women, to the denial of African culture, a classist and political hatred – according to what Grimson (2017) acknowledges as the racialization of politics with the phenomenon of hatred towards *peronismo.* Every contemporary hate speech is an echo augmented by social media and current media of those patterns of national hatred that divide us into whites and blacks – brown people, according to the contemporary vindication – *unitarios* and *federales; peronistas* and *antiperonistas.*

Argentina in pieces as a multimodal play – a critical writing by Piglia and cartoon version of literary works – which tells us about the hate speech topics in the national expression.

> A cartoon of the Argentine violence through fiction. What is that story? The rebuilt of a plot where the traces that the power relationships, forms of violence, leave behind can be retrieved or imagined. Traces in the body and language, first of all, that allow to rebuild the image of the country that writers like. This story must be read holding light to the ´real´ history and its nightmare. (Piglia, 1993: 8)

El Matadero, by Echeverría (2003) is the first literary national text, according to Piglia. In this fiction, traces of the political, classist and racial hatred can be found. The first social cleavage between *unitarios* and *federales* can be strongly perceived in the literary text where the narrator makes use of descriptive language to show the political cruelty of a dictatorship that makes use of violence as the way of ruling and imposing, which under the scope of *unitarios* would be a barbarism.

The social cleavage separates the country from the city, the culture and civilization from the barbarism. It also integrates and involves the population made up of *criollos* and indigenous people together with written documents about black presence, which was almost cleared from the social representations of the Argentine culture.

The literary language of *El Matadero* describes characters that live in a place with a classist, racist and despising perspective. Black women that kill animals are described as monsters, harpies who fight for the flesh. Black African women and *mulatas* are feminine figures of that feast of popular blood where the assassination of a *unitario* is perpetrated in a narrative parallel with the killing of a strong bull.

Hatred generates acts of violence on the part of the tyrant Juan Manuel de Rosas, who through a *mazorca* – term that refers to the governor's paramilitary group and to a corncob, a phallic symbol and instrument of rapes and tortures – threatens citizens and enforces the arbitrary of the institutionalized violence in this police action that the atmosphere imposes. The metaphor of the slaughterhouse with the aim of representing how Argentina will be beating in the collective memory and will be repeatedly expressed in many occasions throughout the history along with *La Resfalosa*[2], which is updated every time the forces of tyrannies appear to prohibit or limit democracy. The Other is the one who does not resist. Its presence, its ideology, its way of being is destroyed in the body with a violence that is applicable to the wretched, to what must disappear. The irony of the display of sliding shows that hatred does not involve compassion because the object of hatred is not a subject, does not belong; it is treated like an obscure object, the insurmountable limit. At the same time, hate speech puts the owners of that violence in a place of barbarism, of animal brutality, of something that cannot be acceptable to build a society. Therefore, hate speech is polarized. Hate speech and acts that do not have any chance of dialogue and are self-excluded. Belonging to a place or another is marked by class, race, religion that maintains hatred and a writing of abjection.

Argentina in pieces considers the Owners of land, by Viñas with a cartoon version made by Breccia (Piglia, 1993: 23-32) and that Piglia chooses to continue with the fictional reconstruction of hatred and violence in the

---

2. Hilario Ascasubi, autor of the poem *La Resfalosa*. It is a poem that narrates the torture of *mazorqueros* to *unitarios*. *La Resfalosa* is also a folklore rhythm, which tone served to execute cruel acts of the Rosas' dictatorship against his political enemies. Hate speech was the dance of the death for the victims that were hanging until they bleed out and "sliding" along their own blood.

establishment of the National State. In this fiction, it can be noticed what the critic calls the oligarchic violence. Turning the dessert conquest into fiction allows to highlight the violence towards indigenous people, objects of violent death, and towards workers. Similar to *El Matadero*, the enemy is animalized. The Other is so dehumanized that it does not even deserve the bullet that kills them. In this way, indigenous people are chased, hunted and assassinated like animals, hauled for fun. Their sexualized bodies are mutilated to show and count deaths, resulting in the production of denial, of destruction of the Other whose lands shall be occupied and civilized.

Illustrated by Solano Lopez, Piglia chooses *Cabecita Negra*, by Rozenmacher (2013: 93-101). A new historical moment that marked the history of hatred and hate speech in Argentina with the rising of *peronismo* and the invention of the phrase *cabecita negra* in his critical analysis where he postulates the idea of "inhabited house". Piglia argues:

> ...*Cabecita Negra* can be considered as an ironic version of *"Casa tomada"*, by Julio Cortázar. Or as a better version of the comment made by Sebrelli to the Cortazar's story: '*Casa tomada* explains that feeling of distress that the invasion of the *cabecita negra* provokes in the middle-class'. Sebrelli's interpretation defines better to Sebrelli instead of Cortazar's story, but, anyway, it has become a common place of criticism and it overlaps with the story. *"Cabecita negra"* is a comment to Sebrelli's comment. Not only because Rosenmacher's text explicitly cited Cortazar's story (´*La casa estaba tomada*´), but also because of the invasion of the private premises of the middle-class by *cabecita negra* who becomes the anecdote of the story. (1993: 91)

That critical reading shows a way of reading the literary series, making the artistic production to be in contact with others that make up the series and critical discourse like the comment by Sebrelli. But it also allows to think that in the hate speech present in the stories by Ronsenmacher, the eyes of the Lord are explicit. Lanari towards his own servant that is treated like an object – his wife took her to their country house – the girl that cries while

being drunken at night and hauls like an animal, and the police officer – another *cabecita negra* – described as a monster, as a walrus.

When in the library of the house invaded by negros, the watcher puts his feet on the table, the historical event of the feet on the font is explicitly mentioned to symbolize the invasion of *cabecitas negras* that was visible on 17th October 1945. Grimson (2017) will rebuild the political uses of the term *cabecita negra* as a way of representing the unknown, the interior, the indigenous, black and poor Argentina that was in the streets to defend their leader Perón. Racialization of politics is almost immediate. Relating *cabecitas* to *peronistas* was almost direct. Even today, the term exists as an expression of structural racism that refers to being together, but separately. According to Piglia's words:

> *Cabecita Negra* can be read as a critical version of that series of texts from *El Matadero*, by Echeverría to *La fiesta del monstruo*, by Borges and Bioy, which represent in a hallucinating way that primitive and brutal world that is present in heads, monsters and fiction characters of the working classes. (1993: 92)

From this historical perspective developed by the literature, we can notice how Argentine history has been pierced by hate speech that symbolizes or represents the Other, who must be destroyed on the basis of racist, classist and xenophobic stereotypes that have lasted in time and that today are also noticeable in contemporary hate speech. Hatred that permeates social media is the same hatred towards indigenous people, who must be killed, depopulated, destroyed in its body, which continues to be an object to conquer. The political cleavage between the oligarchy that exists today is exhibited in many political parties while *peronismo* continues to be the racialization phenomenon of politics, by insisting on the idea that a Peronist is a *planero*, a beneficiary of social security programs, as a result of the Argentine economic disruption. That is why people that must be civilized deserve jail or bullet; therefore, they become objects of repressive violence of the State and of the systematic attacks on social media and mainstream media.

Conclusion

4.1 Departing from the analysis of hate speech and its multiple dimensions, it is postulated that the complex analysis of social discourse allows to adopt a transdisciplinary stance Riveros Argel, Meriño Vergara and Crespo (2020) to study hate speech in a context of social, institutionalized use and as real speech acts. When considering the complexity, different and opposing perspectives come to grips to assess the analyzed phenomenon; for example, between the subjective and legal dimensions as the latter is regarding the tension between freedom of speech and control. However, from the subjective perspective, hatred is thought to be an inevitable force. A death drive, from which nobody can escape, a strong hatred that pushes the human being to malice and cruelty.

The complexity also allows to take a step out of the naïve perspective that opposes hate speech against positive or peaceful speech. On the contrary, the struggle can be understood by the new place of utterances where it is also possible to ponder hate speech as a creative force that seeks to destroy so as to revolutionize and change deep structures in the society.

By adopting a transphrastic perspective on speech, complete series of speech are analyzed. The syntagmatic reading of speech leads to assess the case of online writing, the way online users and prosumers replicate, forward and recreate content. In this algorithmic democracy, technological aspects that influence on the formats and construction of messages and new audiences are analyzed in these new ways of social circulation of information. Radicalization and bias in the contemporary society are constitutive elements of hate speech and its production and circulation. Information bubbles in which each group is locked deeply influences on the construction of the Other as an enemy that must be destroyed.

4.2 What is new about hate speech as a characteristic of the society can be found in the *digital buzz*, that buzz or complaints that becomes hate rhetoric. In the morphology of hatred, what is vague, nomad and viral causes great effect and harm. It gets worse when mixing everything, from the fact of

remembering crimes against humanity that thousands of users post to the attacks and threats to politicians belonging to *peronismo.* We can always find historical roots in the construction of speech that permeates Argentine society as a complex sociocultural patch. Argentina is a racist country even though it is not recognized as such. The problem of racialization of politics is added. It defines us since the origins of the country, in which popular groups are made up of minorities of immigrants coming from Italy, Poland and Spain who constituted the real white and European immigration that Piglia considers as the "polluted Argentine" when analyzing the jargon in *tango* and the stereotype of Italians and Turkish. Moreover, the unwanted Other is made up of *cabecita negra,* indigenous people and *gauchos* from the interior of the country and any other minority that departs from the values of the Argentine oligarchy, criollo by lineage and its ideology.

4.3 In this article, the case of the attempt of assassination of the vice-president Ph.D. Cristina Fernandez de Kirchner evidences how hate speech was materialized in a hate crime that was committed with impunity. Even though the perpetrators of the attempt of magnicide are under arrest, the systematic production of hate speech that was produced in that context or criminal act was not taken into account as part of it. The lack of legislation does not allow to consider shitstorm towards a political group that ended up in a hate crime as a real risk to democracy.

4.4 Another risk to democracy is the supremacy of the algorithm to establish forms of content circulation. The definition of consumers ends up being a lethal weapon against what is common and democratic as it radicalizes political stands to avoid dialogue and polemics, which are the basis of democracy. If each person has an inner conversation in an insurmountable narcissism, it will produce the destruction of the social fabric and will favor neofascist and authoritarian tendencies that in Argentina has been set in the ways to refer to the State through the use of violent and repressive language which led to the victory of the far-right political parties for the following years.

Through collective political actions to support democracy and its values of listening, polemics and debate will be crucial to avoid a reign of hatred in this new political stage. What, in Argentina, has been a public policy around this issue through official departments (Public Defender`s Office for Audiovisual Communication Services, The National Institute against Discrimination, Xenophobia and Racism [INADI], among others) will eventually turn into militancy and intellectual compromise from the groups that are linked to knowledge and expertise. From the great necessity of advancing on the construction of observation groups of hate speech, which apart from the qualitative approach herein explained, could work on the analysis of great amounts of data and incorporate artificial intelligence to monitor and prevent damage of hate speech in social life. So far official organisms and groups that work around this communicational issue of hate speech mostly follow a manual approach like the one followed in this analysis and what is evidenced in contemporary official reports (INADI, 2020, December 10; INADI, 2022, May 17; Ipar, 2020, June).

References

Ámbito. (2023, 27 de junio). El Gobierno nacional denunció penalmente a Gerardo Morales por la represión en Jujuy. *Ámbito.* https://www.ambito.com/politica/el-gobierno-nacional-denuncio-penalmente-gerardo-morales-la-represion-jujuy-n5755574

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info,* (49), 98-124. https://dx.doi.org/10.7764/cdi.49.27817

Angenot, M. (2010). *El discurso social. Los límites históricos de lo pensable y lo decible.* Siglo XXI.

Bilyc, P. (Coord.). (2021, 10 de diciembre). Conversatorio discursos de odio y comunicación violenta. *XXIII Congreso de la Red de Carreras de Comunicación RedCom.* [video] YouTube.

Brizuela, A. (2023, 25 de mayo). Violencia y balas de goma en cortes de ruta de estatales. *Página/12*. https://www.pagina12.com.ar/552531-violencia-y-balas-de-goma-en-cortes-de-ruta-de-estatales

Brizuela, A. (2023, 20 de julio). Jujuy: Cómo fue el brutal ataque a la docente Camila Müller. *Página/12*. https://www.pagina12.com.ar/569618-el-brutal-ataque-a-camila-muller-en-primera-persona

Cebrelli, A. (2022). De resilencias y agencias decolonizadoras. Reflexiones hacia la construcción de saberes articulatorios de género en el campo de la Comunicación. In: A. Sardinha, V. Alves Lima, L. Castro, & V. Belmonte (Ed.), *Decolonialidade, comunicação e cultura* (pp. 158-175). UNIFAP Editora.

CONADEP. (1984). *Nunca más. Informe de la Comisión Nacional sobre la desaparición de personas*. Eudeba.

Congreso de la Nación Argentina. (2006). *Ley 26.150. Programa nacional de educación sexual integral*. https://www.argentina.gob.ar/normativa/nacional/ley-26150-121222/texto

Congreso de la Nación Argentina. (2012). *Ley 26.743. Establécese el derecho a la identidad de género de las personas*. https://www.argentina.gob.ar/normativa/nacional/ley-26743-197860/texto

Congreso de la Nación Argentina. (2015). *Ley 27234. Educar en igualdad: Prevención y erradicación de la violencia de género*. https://www.argentina.gob.ar/normativa/nacional/ley-27234-257439/texto

Congreso de la Nación Argentina. (2019). *Ley 27499. Capacitación obligatoria en género para todas las personas que integran los tres poderes del Estado*. https://www.argentina.gob.ar/normativa/nacional/ley-27499-318666

Congreso de la Nación Argentina. (2021). *Ley 27.636 de Promoción del Acceso al Empleo Formal para personas Travestis, Transexuales y Transgénero Diana Sacayán - Lohana Berkins*. https://www.argentina.gob.ar/normativa/nacional/ley-27636-351815/texto

Corbalán, E. (2022, 17 de julio). Santa Victoria Este: Denuncian la promoción de abusos sexuales en el Trichaco. *Página/12*. https://www.pagina12.com.ar/437585-santa-victoria-este-denuncian-la-promocion-de-abusos-sexuale

Corbalán, E. (2022, 17 de julio). Santa Victoria Este: Denuncian la promoción de abusos sexuales en el Trichaco. *Salta/12*. https://www.pagina12.com.ar/437585-santa-victoria-este-denuncian-la-promocion-de-abusos-sexuale

Defensoría del Público. (2022). *75° Informe de la Defensoría del público sobre discursos de odio*. https://defensadelpublico.gob.ar/75-informe-el-foco-en-los-discursos-de-odio-en-el-marco-del-atentado-a-la-vida-de-la-vicepresidenta/

Di Fátima, B. (2023). A global approach to hate speech on social media. In: B. Di Fátima (Ed.), *Hate speech on social media: A global approach* (pp. 11-15). LabCom Books & EdiPUCE.

Domínguez, A. (2019, 31 de julio). Veganos vs gauchos: ¿Qué hay detrás de la pelea? *Infobae*. https://www.infobae.com/opinion/2019/07/31/veganos-vs-gauchos-que-hay-detras-de-la-pelea/

Echeverría, E. (2003). El matadero. *Biblioteca Virtual Universal*. https://biblioteca.org.ar/libro.php?texto=70300

Galasso, N. (2022, 16 de octubre). 17 de octubre de 1945: Cuando el pueblo decidió andar. *Telam Digital*. https://www.telam.com.ar/notas/202210/607732-17-de-octubre-1945-cuando-el-pueblo-decidio-andar-galasso.html

Grimson, A. (2017). Raza y clase en los orígenes del peronismo: Argentina, 1945. *Desacatos. Revista De Ciencias Sociales*, (55), 110–127. https://doi.org/10.29340/55.1807

Giorgi, G. & Kiffer A. (2020). *Las vueltas del odio. Gestos, escrituras y políticas*. Eterna Cadencia.

Grabosky, S. (2022). Escrituras de odio: Rememoración y actualización de la violencia y represión de estado de la dictadura militar argentina en redes sociales. *Intersecciones En Comunicación*, *1*(16). https://ojsintcom.unicen.edu.ar/index.php/ojs/article/view/142

Grabosky, S. (2023, 31 de julio). El reclamo de los pueblos preexistentes. *Diario Página/12.* https://www.pagina12.com.ar/572786-el-reclamo-de-los-pueblos-preexistentes

INADI. (2020, 10 de diciembre). *Informe Discurso de Odio.* https://www.argentina.gob.ar/sites/default/files/12_01_2021_informe_discurso_de_odio.pdf

INADI. (2022, 17 de mayo). *Empatiz.ar.* https://www.argentina.gob.ar/sites/default/files/2022/05/guia_empatiz.ar_final_accesible.pdf

Ipar, E. (2020, Junio). *Discurso de odio parte 1. Condiciones para su producción y circulación.* Informe LEDA Cualitativo. https://www.unsam.edu.ar/leda/docs/Informe-cualitativo-1.pdf

Kiffer, A. & Donadi, M. (2016). La escritura y el fuera de sí. *Badebec*, *5*(10). https://doi.org/10.35305/b.v5i10.248

Kristeva, J. (1985). *Sobre la abyección. Poderes de la perversión.* Siglo XXI

Lopez Murphy, R. [(@rlopezmurphy]. (2022, 29 de agosto). *Son ellos o nosotros [Tweet] Twitter.* https://twitter.com/rlopezmurphy/status/1564405132871319554?lang=es

Mamaní, M. (2023, 25 de octubre). Condenaron al abogado que cuestionó la reforma en Jujuy. *Página/12.* https://www.pagina12.com.ar/605912-prision-efectiva-para-el-abogado-que-cuestiono-la-reforma-en

Ministerio de Salud. (s.d.). *Plan nacional de prevención del embarazo no intencional en la adolescencia.* https://www.argentina.gob.ar/salud/plan-enia

Miranda, S., Malini, F., Di Fátima, B., & Cruz, J. (2022). I love to hate! The racist hate speech in social media. *Proceedings of the 9th European Conference on Social Media* (pp. 137-145). https://doi.org/10.34190/ecsm.9.1.311

Miró Llinares, F. (2016). Taxonomía de la comunicación violenta y el discurso del odio en Internet. *IDP. Revista de Internet, Derecho y Política*, 22, 93-118.

Oriland [@Oriland1 (2021, 27 de mayo). Gracias @patobullrich. *Twitter.* https://twitter.com/Oriland1

Özlem, A., Gökaliler, E., & Alikılıç, I. (2023). Syrian refugees in the shade of the 'Anti-Syrians' discourse: Exploring discriminatory discursive strategies on Twitter. In: B. Di Fátima (Ed.), *Hate speech on social media: A global approach* (95-119). LabCom Books & EdiPUCE.

Página/12. (2023, 26 de agosto). Récord de reclamos por el tono discriminatorio de entrevista televisiva a matrimonio indígena. *Página/12.* https://www.pagina12.com.ar/582391-record-de-reclamos-por-el-tono-discriminatorio-de-entrevista

Perfil. (2023, 26 de mayo). Salta: Incidentes y 19 detenidos por una protesta salarial de docentes. *Perfil.* https://www.perfil.com/noticias/sociedad/salta-represion-incidentes-y-19-detenidos-por-una-protesta-salarial-de-docentes.phtml

Riveros Argel, P., Meriño Vergara, J., & Crespo, D. (2020). *Documento Número 2. Las diversas definiciones de transdisciplina. Vicerrectoría de Investigación y Desarrollo.* Universidad de Chile. https://doi.org/10.34720/27q7-4s21

Rozenmacher, G. (2013). Cabecita negra. *Obras completas* (pp. 65-71). Biblioteca Nacional

Van Dijck, J. (2016). *La cultura de la conectividad: Una historia crítica de las redes sociales.* Siglo Veintiuno.

Winograd, M. (2020). O ódio na cultura e na clínica psicoanalítica contemporáneas. *Rev. Latinoam. Psicopat. Fund., 23*(4), 724-744. http://dx.doi.org/10.1590/1415-4714.2020v23n4p724.4

Wainfeld, M. (2022, 2 de setiembre). Atentado a Cristina Kirchner: La tragedia no sucedió, los peligros que siguen ahí. *Página/12.* https://www.pagina12.com.ar/478528-la-tragedia-no-sucedio-los-peligros-que-siguen-ahi

Ynofre, A. (2021, 29 de mayo). Chocobar héroe. *Tiktok* [video]. https://www.tiktok.com/@albertoynofre/video/6967880032115100933?q=chocobar%20heroe&t=1691140582373

ACTIVISM AS A RESPONSE TO HATE SPEECH ON SOCIAL MEDIA: IDENTITY AND FANDOM ON THE @PABLLOVITTAR PROFILE

Marcelo Freire
/ Federal University of Ouro Preto, Brazil

Debora Cristina Lopez
/ Federal University of Ouro Preto, Brazil

Daniel Martín-Pena
/ University of Extremadura, Spain

Introduction

At the Lollapalooza music festival in March 2022, Brazilian artist Pabllo Vittar made a political demonstration in support of Luiz Inácio Lula da Silva, at that time a pre-candidate in Brazil's presidential elections this year. The singer's parade, with a banner bearing Lula's face during the performance, led the country's Superior Electoral Court to comply with the request of President Jair Bolsonaro (later defeated in the election) and ban political demonstrations at public events, under penalty of a R$50,000 fine for the organisers. The political demonstration and the court decision generated online debates on issues such as censorship, freedom of expression and Brazil's political future.

In this article we analyse 4,950 comments made on @pabllovittar's Twitter posts, specifically in relation to the Lollapalooza Festival. We developed a qualitative analysis, organised through data coding techniques that guide a deductive approach to the sample.

The comments analysed were made on posts that were not political in nature. Although during Lollapalooza Vittar spoke out, the photograph of this moment was not posted by her on Twitter, but by her[1] followers and fans. The comments were organised into: (a) passionate fan statements to the singer; (b) requests for publication of the political demonstration at Lollapalooza; (c) aggressive demonstrations by Bolsonaro fandom; (d) messages of support for Bolsonaro; (e) transphobic[2] and sexist demonstrations by Bolsonaro voters; (f) misinformation as a strategy to discredit support for the artist; (g) memes, triggered by the two fandoms, reiterating the political position; (h) responses to transphobic and sexist comments with facts, positive images and, above all, praise and "declarations of love" for Pabllo Vittar.

The memes, as a remix of a cultural manifestation, were presented with different strategies. On the one hand, Bolsonaro's supporters bet on the distortion of Pabllo Vittar's political positioning and the appropriation of the "L" made with his hands in support of Lula, resignifying him as a Thief (in Portuguese: Ladrão) and reinforcing the political argument of corruption and misinformation that characterise Bolsonaro's pre-campaign. On the other hand, the singer's fandom opted for humour, with the perspective of the impact of Vittar's demonstration on the Brazilian president at the time.

In this context, we chose to analyze the sample based on the debate about hate speech. The complexity of the concept, especially in the context of our object, allows it to be correlated with fan culture. "Especially at critical times, such as during elections, the concept of hate speech may be prone to manipulation: accusations of fomenting hate speech may be traded among political opponents or used by those in power to curb dissent and criticism"

1. Pabllo Vittar personally identifies as a gay man, but when he is developing his artistic performance she identifies as "she/her". Thus, the authors chose to adopt the identification "she/her" in this chapter.
2. It should be noted that Pabllo Vittar does not identify herself as transgender, but as a gay man performing as a drag queen. Thus, there is no female gender identity, but a fluid one, as she herself presents it. We chose to keep transphobia in the chapter because the attacks identified in the sample often identify him as a transvestite or transsexual and use transphobic discourse.

Activism as a response to hate speech on social media:
identity and fandom on the @pabllovittar profile

(Gagliardone et al., 2015: 10). The country's presidential race, political polarization and Pabllo Vittar's activist character are variables that reveal the country's political context and its actions on digital platforms.

The data shows that positive mobilisation and identification of the fandom with Pabllo Vittar was the main strategy used to combat hate speech in their networks. The cycle of aggressiveness and polarisation characteristic of social media interactions was contained by the fandom's identification with Vittar and by a collectively and naturally constructed strategy to combat hate speech.

Fandom and activism in the connection culture

The audience is the protagonist in the connection society, as defined by Jenkins, Green and Ford (2014). Their involvement in a media production can, according to the authors, define its success or failure. This is not unique to the connection culture. The power of the audience had already been recognised by Jenkins when he presented the convergence culture (2006), which sees in collective intelligence and remix culture crucial points for contemporary communicational practices. This role, we can say, has intensified, resulting in an articulated fandom with an active profile (Masip, 2014). According to the authors, in the culture of connection, fan performance is no longer measured by audience numbers, but by a diversity of variables, such as narrative expansion, online conversations, social networks, consumption of licensed products and integration with content circulating on the Internet. It is a multi-platform scenario, entangling itself in the everyday life of the audience and enabling personalised interactions (Poell, Nieborg, & van Dijck, 2020).

Fans of media products, then, cease to be seen as a subculture, as crazy or exaggerated, and begin to have their place recognised in the media themselves, expanding their interactive possibilities (Jenkins, 2015). Although we cannot say that the public has presented a passive profile in its relationship with the media, Shirky (2011) stresses that technologies,

especially digital ones, have allowed the public to expand its space and have its discourse recognised. In this way, as Lopez and Monteiro Homssi (2021) remind us, the public believes that their opinion or participation, as part of a collective intelligence or representing ideological sharing, matters in the communicative process.

In this way, as Jenkins (2006) reminds us, fans will unite around their common part: the media object. Thus, the "realm of the fan" appears as a space for exchanging information or opinions, reinforcing values and, in the case of our object of research, betting on what Shirky (2011) defines as an exchange of information on a civic level, which seeks to generate transformations in society and is directly related to digital activism.

Understanding hate speech

Matamoros-Fernández and Farkas (2021) present social networks as a space for the proliferation of hate speech, racism and aggression. Citing Sue (2010), the authors point out that discrimination materializes on these platforms through microaggressions or overt discrimination. They argue that this is a reshaping of a "structural oppression based on race, gender, and sexuality as well as their intersectional relationship" (2021: 206). This reshaping is established through the practices of users, but also through the platforms' own affordances and policies, which replicate social practices that are permissive towards those who discriminate, along with the possibility of anonymity and the impunity involved in this process. The platform model is based on interactions and sharing (Hutcher & Helmond, 2018), regardless of their nature, social implications and impacts on guaranteeing human rights.

For Gagliardone et al. (2015), hate speech should be understood as the intersection of multiple tensions. The authors explain that on the one hand it is an expression of social tensions, but that it also represents an understanding of the challenges and opportunities that come with disruptive technologies such as the internet. In addition, they highlight the need to consider and

balance fundamental principles and the guarantee of human dignity in their various practices. Makhortykh and González-Aguilar (2023) explain that although hate speech has multiple definitions, they relate to attacks on specific groups and are therefore characterized as a threat to human rights. This use, normally associated with a justification or incitement to hatred, violence and discrimination, has become more widespread on digital platforms, reaching specific audiences and expanding through sharing between previously isolated users.

According to Gagliardone et al. (2015: 13), although the relationship between online and offline hate speech is close, there are specificities that make digital hate speech more complex, such as its permanence, itinerancy, anonymity and transjuridical nature. Specifically in relation to Twitter, our subject in this chapter, the authors highlight its high capacity for spreading, mainly due to trend topics and the development of threads that enhance incitement to violence. The consequences spread across both digital platforms and offline society. Amores et al. (2021) cite Müller and Schwartz (2020) in revealing a correlation between the increase in online hatred and hate crimes in certain regions. This reality is made even more difficult, as we have said, by the transjurisdictional nature and legislative diversity with regard to these actions.

According to Amores et al. (2021), there has been an increase in online and offline hate speech motivated by ideological reasons, which is what our research is about. In their study, the authors explain that in 2020 the Spanish Ministry of the Interior defined eleven categories to classify crimes against vulnerable populations, namely: (1) racism/xenophobia, (2) political ideology, (3) sexual orientation and gender identity, (4) religious beliefs or practices, (5) disability, (6) gender reasons, (7) antisemitism, (8) aporophobia, (9) antigypsyism, (10) generational discrimination, and (11) discrimination by illness. The first three are indicated as the most recurrent. Of these, two constitute the context of the hate speech we analyzed in the comments on Pabllo Vittar's posts: political ideology and sexual orientation and gender identity.

The coordination of these axes has resulted, as we have seen, in an increase in the aggressiveness and polarization that has characterized Brazilian political relations in recent years. Since the election of Jair Messias Bolsonaro in 2018, the country has experienced a scenario of "systematic production of defamation of government opponents, including journalists; attacks on communication companies and the discrediting of their news; disinformation and fake news about reported happenings and the press being portrayed as an enemy of Brazil" (Capoano, Sousa, & Prates, 2023: 159). As the authors point out, emotion takes center stage in political processes, also triggering the role of individual feelings in this arena, which allows us to understand the functionalities of hate speech in the political process. Emotion and affection are therefore considered to be the driving forces of contemporary society.

Methods

This article conducts a qualitative analysis (Gibbs, 2009) of social media posts to understand how interactions between fans were constructed. The analysis is guided by the perspective of fandom, so it analyses the fan's relationships with other fans, with their nemeses and with the object of idealisation (be it the artist Pabllo Vittar, the Brazilian president at that time Jair Messias Bolsonaro or the former president – now re-elected – Luis Inácio Lula da Silva).

While the object of analysis of this research has many dimensions – political, cultural, identity – in this chapter we look at the fan perspective and how it can be directly or indirectly affected by political standing. We look at it from the perspective of engagement, seeking to meet the following objectives:

· The main objective is to understand how Pabllo Vittar's fandom, which identifies with social politics and the left, responds to the comments of bolsonarist groups (also characterised in this study as fandom).

· It also seeks to observe whether the demonstrations trigger mobilisation and identification with the artist as a strategy of response to hate speech.

Activism as a response to hate speech on social media:
identity and fandom on the @pabllovittar profile

The data for this research was collected manually. The first step was to delimit the observation of comments on posts made by the artist, regardless of their content. The artist's performance at the festival took place on March 25, 2022. We collected comments from three posts, published between March 25 and 27, 2022. At the time, Pabllo Vittar's Twitter profile had more than 3.1 followers. Our sample included a total of 1,490 comments on April 19, 2022.

The posts analyzed were: "@LollapaloozaBr ✨ ✨ ✨ ✨ ✊🏽", with photographic records of the festival; "TODAY HAS COLOMBIA 🇨🇴 LISTOS¿?", announcing a concert she would be performing in that country; and "✋🏽 💕 GRACIAS COLÔMBIA", with a record of the performance.

Based on an exploratory study of a descriptive nature, we sought to understand who were the subjects represented in the media products that involve fandom in the object in question. Based on the understanding that: a) there is ideological sharing; b) there is articulation around the subject, generating actions of collective intelligence and commitment; c) there are narrative expansions in their manifestations or media productions; d) there is an affective bond established between these subjects and the public, we identified three fandoms, established as follows (see Fig. 1):

Figure 1: Fandoms composition in the sample
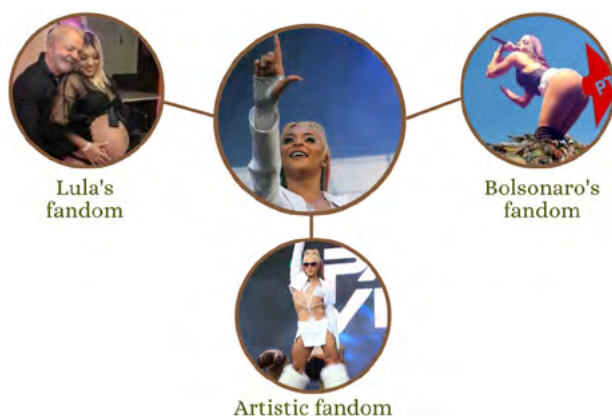
Source: The authors

The Pabllo Vittar fandom identified in the sample of 4,950 comments made on the artist's posts about the Lollapalooza 2022 Festival on her Twitter account is organised into three groups: (a) the eminently artistic fandom, which reiterates its devotion to the artist (as a person or in relation to his work); (b) the Lula fandom, which shares values connected to minority rights and which, to a large extent, form part of the singer's fan group and were triggered by Pabllo's demonstration in favour of the politician at the concert in question; (c) Bolsonaro's fandom, also triggered by Pabllo's demonstration, but which is not necessarily part of the artist's fan group, but appears in her network to respond to her political positioning. In the sample analyzed, these fandoms accounted for 62.79% (artistic fandom), 23.68% (Lula's fandom) and 12.90% (Bolsonaro's fandom).

In the research protocol adopted, we sought to code the data in search of patterns that would allow us to compare the records and recode them, analysing them on the basis of these patterns (Yin, 2016). We adopted the open coding model, with conceptualisation leading to classification (Strauss, 2008: 105), with fandom as the starting concept and hate speech as the concept of analysis generated by the data. We opted for the descriptive mode of interpretation, with a narrative presentation of the data (Yin, 2016).

The conceptualisation-oriented protocol was applied to the 4,950 comments on the four posts about the Lollapa-looza 2022 Festival on Pabllo Vittar's Twitter account. The data, collected on 19 April 2022, is presented and analysed below.

Results

Pabllo Vittar is a Brazilian singer who is part of a new generation of artists that Chris Malone Méndez called in Forbes magazine the "second Latin boom". Self-identified as gay and unproblematic with her gender identity (Essinger, 2017), she was named by Forbes magazine as one of the leaders of the new generation (Chow, 2019). Her first hit, the video "Open Bar", was produced in 2015 at a friend's house and cost a total of 600 reais. From this point on, career advancement was relatively fast.

> While she may not be the first Latin American artist to release music in different languages, she is certainly the first artist of her kind to do so: in just three years, Pabllo Vittar has built her own musical and cultural empire that transcends borders and boxes, uniting music fans across the Americas and arguably becoming the world's most popular drag queen in the process (Malone Méndez, 2020: online).

Pabllo Vittar's music mixes electronic and Brazilian rhythms such as tech-nomelody, arrocha, Carioca funk and forró. It has a distinctly Brazilian musical identity, but sounds closer to styles such as bachata and reggaeton.

Vittar is known for his activism for LGBTQIA+ rights, including political demonstrations involving the issue. In the last Brazilian election period, in 2018, she was an active voice in the resistance movement to Jair Messias Bolsonaro, known for his sexist and homophobic actions and speeches (Lopes & Castro, 2022; Guazina, Guerreiro Leite, & Santos, 2021; Maranhão Filho, Coelho, & Dias, 2018).

Her actions and demonstrations reveal the defence of human rights and go against Bolsonaro, even in her profession. In addition to positioning herself on social networks and in concerts, the singer has already broken her contract with a shoe company because of her relationship with Bolsonaro (Pabllo Vittar..., 2018). After ending her relationship with the shoe brand Vincenzza, the artist posted on her social networks:

> "From the beginning of my career, I always knew it would be very difficult to get support from brands that wanted to engage with the LG-BTQIA+ drag artist that I am. Many doors have closed, but some have opened and with that I have worked so far with partners for which I am very grateful. I leave here my thanks for the support so far, but I couldn't reconcile my work with a discourse that makes it clear that I don't care about the human rights of the entire LGBTQIA+ community, of which I am a part", she wrote in a series of stories on Instagram. (Vittar, 2018: online)[3]

3. https://exame.com/marketing/pabllo-vittar-rompe-com-marca-de-sapatos-por-causa-de-jair-bolsonaro/

The singer's social networks, which have already received awards in Brazil, Latin America and Europe, are far-reaching. As of April 2022, 3.1 million people on Twitter, 12.5 million on Instagram, 2.2 million on Facebook, 7.8 million on TikTok, plus 5 million listens per month on Spotify and 1.4 million views on her YouTube channel.

Pabllo Vittar's activism represents a threatened group in Brazil. According to a survey by Grupo Gay da Bahia (2022), the country had 135 LGBT+ violent deaths in the first six months of the year, an average of 22.5 per month. According to the report (2022: 1), "gays appear as the biggest victims (46.66%) compared to 42.96% of transvestites, transsexuals and trans women".

Figure 2: Photograph of Pabllo Vittar at Lollapalooza 2022

Source: https://cutt.ly/t0OCSb1

This was the context of the Lollapalooza Brazil Festival, held between 25 and 27 March 2022. In her performance on the first day of the event, Pabllo Vittar shouted "Out with Bolsonaro" and walked through the audience with a towel representing Lula's face (Fig. 2). The demonstrations follow the singer's political alignment and activist profile and generated repercussions both on social networks and at the event itself.

Among the reactions was the decision of Brazil's Superior Electoral Court (TSE), published on Saturday 26th, which prevented political demonstrations by artists during the festival. According to the TSE, the Lollapalooza

Activism as a response to hate speech on social media:
identity and fandom on the @pabllovittar profile

organisers would have to pay a fine of R$50,000 in case of non-compliance. The court claims that the demonstrations constituted advance electoral propaganda for the 2022 presidential elections, in line with complaints made by Jair Bolsonaro's Liberal Party (Amorim, 2022).

The decision, considered censorship by journalists, artists and academics (Sakamoto, 2022), generated a wave of new demonstrations by artists at the festival and other events. Among them are Marina, Emicida, Miley Cyrus, Jão, Detonautas and Anitta (Amorim, 2022).

On the artist's social networks, the manifestations have multiplied. The first pattern we observed emerged in the comment interaction flows and refers to the place occupied by Pabllo. The singer does not respond to the comments. Her action is reduced to the posts itselfs, even if they are made up of calls to the public, with invitations to action or posts of an emotional nature, thanking fans, talking about the excitement at events and even in advertising a show inviting fans to join in the experience (Fig. 3).

Figure 3: Interaction flows in the analysed comments

Source: The authors

In the flow of interactions, few were bilateral. This occurred in the interactions between the fandoms. Pabllo Vittar's fans had as their main posting pattern the reiteration of the link with the singer (using the expression "mama", talking about her experience in the show, praising her performance). In the other hand, Lula's fans reinforced the political link with her and her colleagues. Some strategies were praising her public position in the democratic process, warning her about the persecution and threats, calling for new political demonstrations, defending her from attacks and responding to the posts of the bolsonarista fandom. Lula's fandom also called for new political demonstrations, defending her from attacks and responding to posts by Bolsonarista fandom.

Among the posts from Lula's fandom, 15.18% called for new public demonstrations of support, 11.61% declared support and were willing to defend the artist, 16.96% featured Lollapalooza records linked to Pabllo's political demonstration, 11.61% featured declarations of love for the artist and Lula, 17.86% registered support for Lula or made an explicit political campaign and 8.93% responded to Bolsonaro attacks or produced original attacks, many of them linked to the president's health.

Meanwhile Bolsonaro's fans attacked the artist, Lula and his voters had lighter approaches, with some exceptions that had used non-informative content, created aggressive memes, unleashing homophobic and transphobic speeches and campaigned for Bolsonaro in the 2022 presidential elections.

The singer was the recipient of many messages, whether of affection (a); political identification (b); request to expand political demonstrations and share militancy (c); of hate speech and aggressive fanaticism from those who do not share Pabllo Vittar's values (d); of misinformation and malicious content, always ignored by the artist in her positioning on the networks (e); of questioning about the demonstration (f), as exemplified below.

a. (a) "The level of beauty of this drag is inhuman" / "Flawless beauty! I love that she has a round face, she was super cute" / "You look cool

mommyyyyyyy" / "Thank you for being the greatest in everything you set out to do. I LOVE YOU" / "Perfection has a name, a surname, a trajectory and international demand".

b. (b) "An artist who is an artist takes a stand! Thank you, Pabllo, for being our voice. We are waiting for you to be very diva at Lula's inauguration!". / "Pabllo is wonderful and represents me too much at this moment! and long live Lulapalooza 2022" / "Daddy Lula always present".

c. (c) "Can Lula's flag fly in Colombia mommy? It's going to be dressed in red, give us a spoiler". / "Mommy publishes photo with Lula's towel" / "Will the @TSEjusbr punish the public holding the towel with @LulaOfficial's face on it? Hahaha. #LULApalooza #lollapaloozaBR".

d. (d) "She'll never be a woman... the trunk is hidden there" / "Festival Rola Pro Lula (Trunk to Lula)" / "And the eggs accompany Lula wherever he goes" / "First of all, stop swallowing helium gas to speak, because singing... Hey, you'll need thousands of incarnations to start in the shower".

e. (e) "Imagine how much Lula did not pay, and with taxpayers' money. They say that, if Lula wins, Pablo will be Minister of Education" / "Look, that's what this rouanet law is for, for the artist to manipulate people. Many artists went crazy without rouanet".

f. (f) "Is the event about music or politics? hahahahahaha".

Discussion

The initial proposal of this study was to understand how the Pabllo Vittar fandom reacted to the demonstration through an analysis of the comments on social networks. However, after conducting an exploratory descriptive study, we noticed the presence of three different fandoms, as already presented in the methodology. We then chose to reorganise the analysis into eight categories that consider the fandom of origin and the different places of speech as variables in the deductive analysis of the sample.

The comments were organised into these eight categories: (a) passionate fan statements to the singer (62.79%); (b) requests for publication of the political demonstration at Lollapalooza (3.59%); (c) aggressive demonstrations by Bolsonaro fandom (8.42%); (d) messages of support for Bolsonaro; (e) transphobic and sexist demonstrations by Bolsonaro voters (5.92%); (f) misinformation as a strategy to discredit support for the artist (5.07%); (g) memes, triggered by the two fandoms, reiterating the political position (3.17%); (h) responses to transphobic and sexist comments with facts, positive images and, above all, praise and "declarations of love" for Pabllo Vittar (22.3%). The sum of the values exceeds 100% because some posts were classified in more than one category, as it was a complex phenomenon that triggered more than one strategy.

Passionate manifestations and declarations of love from Pabllo Vittar's fans are predominant in the sample and also interact with other categories, mainly with Lula's fans, who declare their love for both. The passion and affective bond that characterise a fan and lead them to interact and follow in the artist's footsteps (Jenkins, 2015) predominate in the comments analysed. And even without Pabllo responding to the fans in the publications, the creation of the bond occurs through the constructed universe (Lopez & Monteiro Homssi, 2021), that is: performances in shows, albums and video clips, public manifestations in social networks and interviews, activism and defence of LGBTQIA+ rights and publications on social networks, always close to the audience, in the form of dialogue and which trigger the emotionality of those who accompany her.

In this context, the compliments become a form of recognition of the artist's place and her ability to speak to her fans. They are compliments to the character created for the public appearance and which reiterate the collective existence of these subjects who fight for the right to a life free from violence.

The singer's activist profile is reflected in publications calling for the circulation and reproduction of photos of the political demonstration during Lollapalooza. The affiliation with Lula, in the sensitive and critical political

moment Brazil is going through, with threats to the democratic state of law (Estudo: Brasil..., 2022; Chade, 2022; Mena, 2022), marks the struggle for the right to exist of an oppressed group and for the right to be a democracy. The national context expands the mobilising potential of public demonstrations, combining artistic performance and political action (Fig. 4).

Figure 4: Pabllo and Lula's fandoms integration commentary

Source: https://twitter.com/pabllovittar

The fan production, seen in the figure above, represents a high level of approximation and identification with the artist. Her connection with a personal analysis, the "proud of you", with the association with the courage to fight for what one believes in a context of attacks on democracy and the appropriation of the expression that grew in the networks, Lulapalooza, reveals the potential for collective articulation that characterises digital interactions and the potential for political mobilisation that exists in fandom.

The hashtag #Lulapalooza was appropriated by both fandoms: on the one hand, Lula's supporters used it as a symbol of the festival's achievement as a reaction to the censorship imposed on artists. On the other hand, we identify the use of the expression Lulapalooza to spread fake news. The meaning was connected to the claiming that the demonstrations were articulated by the former Brazilian president and the Workers' Party as a strategy to reach out to young voters or even to spread allegations of diversions of left-wing funds in the field of culture. The fake news disseminated by the bolsonarist groups (and many of them replicated in the comments of the sample) were not exclusively produced by the public, but included media aligned with Bolsonaro's government, as is the case of radio Jovem Pan (Copolla, 2022).

Fake news was not the only strategy used by Jair Bolsonaro's fandom in the comments analysed. In a similar proportion to the disinformative links and memes (5.92%), aggressive, threatening, homophobic and transphobic messages (5.07%) and images were sent mainly to Pabllo Vittar. The incitement to violence (Fig. 5) present in these comments follows the profile of the militaristic, militaristic, aggressive and misogynist discourse that has characterised Bolsonaro and his supporters since before his election to the Brazilian presidency.

Figure 5: Aggressive and homophobic comments from bolsonarista fandom

Source: https://twitter.com/pabllovittar

In the comments there is a constant homophobic approach that ridicules Pabllo Vittar for being gay, that attacks and devalues her. The attack on her gender identity guides the Bolsonaro fandom's demonstrations. Reactions to these comments come from the other two fandoms, but with different strategies. On the one hand, the messages are ignored (the homophobic ones, directed at Pabllo, and the homophobic and accusatory ones, directed at Lula) – which generates a cascade reaction from the Bolsonaristas, with disinformation, ridiculing comments or inciting violence, generating a flood[4] of messages.

Specifically in relation to the responses, the fandom of the former PT president used support (with the creation of memes connecting the images of Pabllo and Lula) as a central strategy in their comments. In these posts, emotion was the narrative engine. Thus, political identification and support for Lula, considered part of the struggle for human rights in Brazil, play a leading role in the display of this fandom. In both groups, messages of support for the candidates in the presidential election emerged.

Other strategies were also used less frequently by Lula supporters, such as discrediting the messages sent by bolsonaristas. One example is the evaluation of the profiles of users who posted the hashtag #SouJovemSouBolsonaro (#ImYoungImBolsonaro), discussing their suitability for the category "young". Throughout the sample we found an aggressive manifestation, which we can frame as hate speech, coming from the left-wing fandom. In a photomontage posted twice in the comments, Bolsonaro appears in a hospital bed and Pabllo Vittar in front, with a joyful expression. The user asks, "Pabllo, how do you feel about making the old man feel sick?"

Conclusions

The informational context was fundamental to understanding the phenomenon studied in this chapter. Starting from the conception of hate speech as

---

4. Flooding occurs when one or more people send a large number of messages of interest to a smaller number of people.

an attack on diversity and dignity, it is necessary to understand the place where individuals speak, the socio-political-cultural scenario in which the actors are inserted and their capacity for collective articulation. These are variables that affect each other, generating greater or lesser involvement and identification with the collective actions of fandom.

Understanding this context – and discussing the complexity of political relations and the Brazilian democratic scenario in 2022 – would only be possible through a case study. For this reason, we have chosen to analyse, from a multidimensional and multimethod perspective, the comments of one of the cases of censorship of individual demonstrations in the country. We know that the Pabllo Vittar case cannot be generalised to the reality of political-cultural demonstrations in Brazil (Goldenberg, 2004), but comprehending it allows us to observe the exchanges and understand some of the communicative strategies that are designed for the presidential elections in the country.

The analysis carried out revealed predominant profiles of each of the three fandoms: 1) the artistic fandom based its comments on affection, affective relationships and the attempt to get closer to an artist who represents the values of her audience; 2) Lula's fandom used emotion as a main strategy – whether linked to the struggle for democracy, or linked to support for the PT in the presidential elections – but always relying on ideological sharing with the artist; 3) Bolsonaro's fandom unleashed two strategies (the circulation of fake news and homophobic and aggressive discourse) that had a common goal: To discredit Pabllo Vittar and, in doing so, to discredit Lula, who had his images and places of expression linked by the singer's position-taking.

As for the interactions and responses between the actors in this communicative process, we note a certain one-sidedness, as only the fandoms engage in dialogue with each other and address speeches to Pabllo, who does not respond. The exchanges occurred mainly between the Bolsonaro and Lula fandoms, with different strategies. The first one was dominated by attempted mockery and aggression. In the second, affective narrative, defence of

Lula and, to a lesser extent, aggression. We highlight what we saw as one of their main strategies to relate to the attacks on Lula: the non-response, which seemed to affect especially the bolsonarista fandom, which initiated processes of disinformation "flooding" in the comments.

Acknowledgements/Support

References

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando o ódio ideológico no Twitter. Desenvolvimento e avaliação de um detector de discurso de ódio por ideologia política no Twitter em español. *Cuadernos.info*, (49), 98-124. https://dx.doi.org/10.7764/cdi.49.27817

Amorim, D. (2022, 27 de março). Após 'Fora Bolsonaro' de Pabllo Vittar no Lollapalooza, TSE veta atos no evento. *UOL*. https://noticiasdatv.uol.com.br/noticia/celebridades/apos-fora-bolsonaro-de-pabllo-vittar-no-lollapalooza-tse-veta-atos-no-evento-77939

Bucher, T. & Helmond, A. (2018). The affordances of social media platforms. *The SAGE handbook of social media*, 1, 233-253.

Capoano, E., Sousa, V. de, & Prates, V. (2023). Circulation systems, emotions, and presenteeism: Three views on hate speech based on attacks on journalists in Brazil. In: B. Di Fátima (Ed.), *Hate speech on social media: A global approach* (pp. 159-184). LabCom Books & EdiPUCE.

Chade, J. (2022, 11 de maio). *ONU recebe denúncias sobre ataques do governo contra eleição e democracia. UOL*. https://noticias.uol.com.br/colunas/jamil-chade/2022/05/11/onu-recebe-denuncias-sobre-ataques-do-governo-contra-eleicao-e-democracia.htm

Chow, A. R. (2019, 10 de outubro). *How this Brazilian drag queen is taking the pop world by storm - And fighting for LGBTQ rights along the way. Forbes.* https://time.com/collection/next-generation-leaders/5692972/pabllo-vittar-next-generation-leaders/

Copolla, C. (2022, 28 de março). Boletim Copolla: LulaPalooza,D-2 e as verbas bilionárias para a 'cultura' no Congresso. *Jovem Pan.* https://jovempan.com.br/videos/programas/boletim-coppolla/lulapalooza-d2-e-as-verbas-bilionarias-para-a-cultura-no-congresso.html

Essinger, S. (2017, 27 de janeiro). *Pabllo Vittar: 'Pra dar close, a gente vai até num vulcão em erupção'. O Globo.* https://oglobo.globo.com/cultura/pabllo-vittar-pra-dar-close-gente-vai-ate-num-vulcao-em-erupcao-20830881

Estudo: Brasil é exemplo de piora global da democracia. *Deutsche Welle*, 23 feb. 2022. https://www.dw.com/pt-br/estudo-aponta-brasil-como-exemplo-de-piora-global-do-estado-da-democracia/a-60890702

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech.* Unesco Publishing.

Gibbs, G. (2009). *Análise de dados qualitativos.* Artmed.

Goldenberg, M. (2004). *A arte de pesquisar: Como fazer pesquisa qualitativa em Ciências Sociais.* Record.

Grupo Gay da Bahia. (2022). *Informe Dia do Orgulho Gay, a violência como pauta.* https://static.poder360.com.br/2022/06/pesquisa-grupogaybahia-28jun2022.pdf

Guazina, L. S., Guerreiro Leite, G., & Santos, É. (2021). A normalização da agenda anti-gênero de Jair Bolsonaro: Uma análise dos jornais Folha de S. Paulo e Estado de São Paulo. *Sur Le Journalisme, About Journalism, Sobre Jornalismo, 10*(1), 44–61. https://doi.org/10.25200/SLJ.v10.n1.2021.453

Jenkins, H., Green, J., & Ford, S. (2014). *Cultura da conexão. Criando valor e significado por meio da mídia propagável.* Aleph.

Jenkins, H. (2009). *Cultura da convergência.* Aleph.

Jenkins, H. (2015). *Invasores do texto: Fãs e cultura participativa.* Marsupial Editora.

Lopes, M. M. & Castro, R. P. de. (2022). Discurso de ódio homofóbico nas redes sociais: Uma análise a partir das falas públicas da família Bolsonaro. *Revista Periódicus, 1*(17), 118–139. https://doi.org/10.9771/peri.v1i17.37635

Lopez, D. C. & Monteiro Homssi, A. (2021). Cuando se encuentran cultura fan y podcasting: Las nuevas relaciones con la audiencia en O Caso Evandro. *Hipertext.net,* 23, 93- 103. https://doi.org/10.31009/hipertext.net.2021.i23.09

Makhortykh, M. & González-Aguilar, J. M. (2023). Is it fine? Internet memes and hate speech on Telegram in relation to Russia's war in Ukraine. In: B. Di Fátima (Ed.), *Hate speech on social media: A global approach* (pp. 75-94). LabCom Books & EdiPUCE.

Malone Méndez, C. (2020, 30 de abril). Pabllo Vittar's multilingual music is, above all, a gift to her fans. *Forbes.* https://www.forbes.com/sites/cmalone/2020/04/30/pabllo-vittars-multilingual-music-is-above-all-a-gift-to-her-fans/#21e6295f76fe

Maranhão Filho, E. M. de A., Coelho, F. M. F., & Dias, T. B. (2018). Fake news acima de tudo, fake news acima de todos: Bolsonaro e o kit gay, ideologia de gênero e fim da família tradicional. *Revista Correlatio, 2*(17), 118–139. https://doi.org/10.9771/peri.v1i17.37635

Masip, P. (2014). Audiencias activas, democracia y algoritmos. *Anuario ThinkEPI*, 8, 260-263.

Matamoros-Fernández, A. & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television and New Media*, 22(2), 205–224. https://doi.org/10.1177/1527476420982230

Mena, F. (2022, 13 de janeiro). Ataques à democracia no Brasil se somam a velhos problemas de direitos humanos, diz ONG. *Folha de S. Paulo.* https://www1.folha.uol.com.br/poder/2022/01/ataques-a-democracia-no-brasil-se-somam-a-velhos-problemas-de-direitos-humanos-diz-ong.shtml

Rocha, M. (2022, 28 de junho). Sob governo Bolsonaro, pessoas com licença para armas de fogo disparam e crescem 473%. *Folha de S. Paulo*. https://www1.folha.uol.com.br/cotidiano/2022/06/sob-governo-bolsonaro-acesso-a-armas-cresce-39-mas-apreensoes-caem.shtml

Sakamoto, L. (2022, 29 de março). Censura ao Lolla: Em 4 anos, bolsonarismo foi do desprezo ao medo de Pabllo. *UOL*. https://noticias.uol.com.br/colunas/leonardo-sakamoto/2022/03/29/censura-ao-lolla-em-4-anos-bolsonarismo-foi-do-desprezo-ao-medo-de-pabllo.htm

Shirky, C. (2011). *A cultura da participação: Criatividade e generosidade no mundo conectado*. Zahar.

Strauss, A. (2008). *Pesquisa qualitativa: Técnicas e procedimentos para o desenvolvimento de teoria fundamentada*. Artmed.

Vittar beneficiada pela Lei Rouanet? Veja essa e outras 5 notícias falsas sobre a cantora. *GaúchaZH*. 27 dec. 2017. https://gauchazh.clicrbs.com.br/cultura-e-lazer/musica/noticia/2017/12/pabllo-vittar-beneficiada-pela-lei-rouanet-veja-essa-e-outras-5-noticias-falsas-sobre-a-cantora-cjboyn1dh03f301p91whjzix1.html

Vittar rompe com marca de sapatos por causa de Jair Bolsonaro. *Revista Exame*. 03 sept. 2018. https://exame.com/marketing/pabllo-vittar-rompe-com-marca-de-sapatos-por-causa-de-jair-bolsonaro/

Yin, R. K. (2016). *Pesquisa qualitativa: Do início ao fim*. Penso.

Activism as a response to hate speech on social media:
identity and fandom on the @pabllovittar profile

POLITICS, JOURNALISM AND HATE SPEECH IN PAKISTAN

Imran ul Haq
/ Bahria University, Pakistan

Shabir Hussain
/ Bahria University, Pakistan

Jamal ud Din
/ University of Swat, Pakistan

Introduction

For working journalists in Pakistan, the incidents of trolling, harassment and hate speech has seen an alarmingly rising trend within the past decade. Journalists in Pakistan encounter substantial trolling from supporters of diverse political factions. This matter gained prominence in 2020 when more than a hundred female journalists filed a petition citing online harassment from social media users affiliated with the populist political party, Pakistan Tehreek Insaaf. Despite deliberations in the nation's parliament and concerns voiced by media oversight bodies, no actions have been taken to improve the situation, as observed by Hussain, Bostan, and Qaisarani (2022).

According to the Coalition for Women in Journalism, Pakistan is among the top five worst offenders globally regarding online harassment against journalists. Additionally, the Human Rights Watch report (2022) portrayed the pervasive atmosphere of fear among journalists in Pakistan, highlighting recurrent death threats and harassment on digital platforms. According

to reports from the media, the Pakistan Tehreek-e-Insaaf government employed social media teams to counteract critical journalists, particularly targeting female journalists who were subjected to excessive online abuse, as documented by Mughal (2022). Various sources have indicated that hashtags and trends were orchestrated against female journalists to deter them from providing critical coverage of government policies, as reported by Dawn (2020).

Numerous media watchdogs have associated this issue with the policies of the PTI government aimed at curbing media freedom within the country, as highlighted in the Human Rights Watch World Report (2021), the International Media Support (2020), and Reporters without Borders Report (2021). Reporters without Borders, in their 2021 annual report, included PTI leader Imran Khan in the list of those who suppress press freedom, cautioning against the growing religious conservatism and populism posing threats to media freedom and vulnerable groups such as minorities and women. The report lamented the regulatory bodies governing electronic and social media, suggesting they have become subservient to government agendas, leading to a shrinking space for dissent and freedom of expression within the nation.

Additionally, the Human Rights Watch World Report of 2021 asserted that alongside severe restrictions on traditional media, the government actively engages in online harassment of journalists to impede their critical reporting efforts.

There is no denying the fact that suppressing freedom of expression and curbing the voice of journalists by governments, political parties and their affiliates is a condemnable act in all its forms and manifestations. But still the larger debate surrounding the role of press and journalists in a democracy as forth estate and watchdogs, necessitates a holistic investigation into the working modalities of journalists in a semi-democratic and multiparty political environment of Pakistani politics. For this purpose, we need to consider gradual political alignment of media groups and journalists since

2013 that witnessed a score-setting approach of TV talk show anchors and a clear political tilt of large media groups like Geo News and ARY News. Simultaneously, we also need to understand what the Pakistani public generally expected from journalists in terms of their reporting toward the leadership and performance of governments, opposition and political parties. We argue that in contrast to the well-established democracies, well--defined role of press and journalists, literacy ratio and political maturity of masses in the Western world, these factors tend to be developing or loosely--defined in the political system of Pakistan and hence need to be considered while analyzing hate speech trends against journalists in Pakistan.

It is, therefore, pertinent to mention here that this work does not simply embark upon tracing varieties of hate speech along with their quantitative and qualitative patterns in the Pakistani society rather we bring forth multidimensional causes leading to hate speech against journalists in a politically polarized environment by drawing upon journalistic norms, ethics and social responsibility theory to highlight that how political differences, social inequalities, economical deprivations may aggravate masses to resort to the worst degree of expression generally known as hate speech.

To present such an analysis, this chapter provides an account of social media uses and users in Pakistan with a special emphasize on how two major political parties including the Pakistan Tehreek-e-Insaaf and the Pakistan Muslim League Nawaz utilized social media platforms to mobilize support. Next, we present findings of two case studies related to online trolling and harassing of political and female journalists in Pakistan. Later, we assess the extreme polarized political environment and instances wherein journalists took sides in favor of PTI or Pakistan Muslim League Nawaz. Further, we examine the post-PTI government plight of pro-PTI journalists and their allegations for being fed with corruption content by spy agencies against the leadership of Muslim League Nawaz. In the conclusion section, we extend and superimpose this scenario upon the principles of journalistic norms and social responsibility theory to examine how journalists adopted diverged reporting patterns which collided with public expectations and sentiments.

Internet and social media landscape in Pakistan

To understand the phenomenon of social media hate speech against journalists, government institutions or any other segment of society in Pakistan, we need to take a brief overview of social media penetration in the Pakistani society. According to data from the Pakistan Telecommunication Authority (PTA), Pakistan had surpassed 100 million broadband subscribers and exceeded 50 million internet users by 2020. Furthermore, the country boasted approximately 37 million active social media users during the same period, with Facebook emerging as the most widely used platform, followed closely by YouTube and WhatsApp. Concurrently, the State Bank of Pakistan noted a notable surge in digital transactions, with the first quarter of the fiscal year 2020-21 witnessing a total of 32.9 million e-banking transactions, amounting to Rs 8.1 trillion. This data underscores the escalating significance of social media platforms in facilitating e-commerce endeavors across Pakistan.

Research undertaken by the Pakistan Institute of Legislative Development and Transparency (PILDAT) revealed that more than 30% of Pakistan's population utilizes social media as a platform for political involvement. This involvement encompasses activities such as discussing political matters, disseminating political content, and engaging in political campaigns. These findings offer insight into the extensive uptake and diverse influence of social media within Pakistani society, underscoring its significance across a range of sectors. Nonetheless, it is imperative to complement these quantitative findings with qualitative research and firsthand observations to cultivate a comprehensive comprehension of social media dynamics within the Pakistani milieu.

Social media platforms - Early adoption and growth

Social media channels such as Facebook and Twitter experienced a surge in popularity in Pakistan during the late 2000s and early 2010s. As per a report by the Digital Rights Foundation (DRF) Pakistan in 2013, Facebook

emerged as the most favored social networking platform in the nation, boasting over 9 million users (Digital Rights Foundation, 2013). These statistics suggest a significant societal shift in Pakistan, transitioning from a more conventional way of life to one characterized by a fast-paced and dynamic lifestyle influenced by social media. Consequently, providing a concise overview of this transformation from diverse perspectives would facilitate a deeper understanding of this evolving landscape.

Social media platforms have emerged as pivotal forces in shaping political discourse and rallying the masses in Pakistan. Numerous studies have underscored the pivotal role played by platforms such as Facebook, Twitter, and WhatsApp in political campaigns, activism, and the dissemination of awareness. According to a research report by Repucci (2020), the Pakistani government has increasingly imposed restrictions on online content and targeted social media users who voice criticism against the government, underscoring the profound significance of social media as a conduit for political expression.

Social media platforms additionally function as hubs for small businesses and entrepreneurs within Pakistan. Research delves into the role of platforms such as Facebook and Instagram in facilitating e-commerce and digital marketing endeavors. As per data from the Pakistan Telecommunication Authority (PTA), the country boasted over 50 million internet users in 2020, underscoring the extensive potential of social media for fostering economic activities. Literature also addresses the array of challenges and opportunities associated with social media utilization in Pakistan. These encompass concerns regarding privacy, cyber security, misinformation, and the digital divide. Organizations like the Digital Rights Foundation and Bytes for All Pakistan have conducted research and advocacy efforts centered on digital rights and online freedoms within Pakistan, emphasizing the imperative for policy interventions and initiatives aimed at enhancing digital literacy.

Hate speech on digital platforms in Pakistan

Social media platforms serve as convenient and often anonymous channels for both individuals and groups to disseminate hate speech. Research conducted by Saleem and Rehman (2017) sheds light on how platforms like Facebook, Twitter, and WhatsApp are utilized to propagate divisive narratives rooted in ethnicity, religion, and ideology within Pakistan. These platforms facilitate the rapid dissemination of hate speech, exacerbating social tensions and conflicts. Moreover, social media algorithms have the potential to amplify extremist viewpoints by prioritizing content that aligns with specific ideologies or biases. Tufail, Nawaz, and Rana (2018) explore how echo chambers and filter bubbles on social media platforms reinforce polarized perspectives and foster the proliferation of hate speech in Pakistan. Consequently, users may find themselves increasingly exposed to extreme content, potentially leading to radicalization and the perpetuation of intolerance.

One potential factor allowing groups to exploit these platforms for hate speech is the often inadequate content moderation on social media sites, permitting the unchecked proliferation of hateful content. Hussain and Ali (2020) delve into the hurdles encountered by social media companies in regulating hate speech within Pakistan, citing challenges related to linguistic diversity, cultural subtleties, and limited resources. Ineffectual enforcement of community standards and the inconsistent application of content policies further exacerbate the persistence of hate speech online. Consequently, hate speech on social media undermines social cohesion and exacerbates tensions between different groups in Pakistan. Research conducted by the Digital Rights Foundation (DRF) underscores how hate speech directed at religious and ethnic minorities contributes to their marginalization and discrimination. The normalization of hate speech online may foster an atmosphere of fear and hostility, impeding progress towards pluralism and inclusivity.

Similarly, political entities and extremist organizations frequently utilize social media platforms to foment hatred and garner backing for their causes. PILDAT's (2019) examination of democracy and political parties in Pakistan delves into the convergence of political polarization and identity politics with social media dynamics, resulting in the escalation of hate speech during electoral campaigns and political gatherings.

Studies examining the impact of social media on political polarization within Pakistani society indicate that these platforms have emerged as crucial spaces for the articulation and consolidation of political rifts. The algorithms employed by social media platforms typically display content that resonates with users' preexisting beliefs and inclinations, fostering the creation of echo chambers and filter bubbles. This trend restricts exposure to diverse viewpoints and reinforces prevailing biases, thereby intensifying political polarization. Tufail, Nawaz, and Rana (2018) delve into how social media users in Pakistan frequently interact with individuals who share similar perspectives, resulting in the amplification of polarized sentiments.

Social media platforms serve as breeding grounds for misinformation and disinformation, exacerbating political polarization through the dissemination of false or deceptive narratives. Research conducted by entities such as the Digital Rights Foundation (DRF) and the Pakistan Institute of Legislative Development and Transparency (PILDAT) has documented instances of misinformation campaigns and the proliferation of fake news aimed at manipulating public opinion and deepening political rifts. Furthermore, social media platforms are extensively leveraged for political campaigning and activism in Pakistan. While these endeavors may galvanize supporters and enhance awareness, they also foster polarization by engendering a dichotomous "us-versus-them" dynamic and nurturing animosity between opposing political factions. Hussain and Ali (2020) investigate that how social media has been wielded by political parties and interest groups to polarize public sentiment and influence electoral outcomes. Consequently, social media facilitates the propagation of polarizing rhetoric and hate speech, particularly

during politically charged events or debates. Scholars like Saleem and Rehman (2017) have scrutinized the utilization of social media platforms to propagate divisive narratives centered on ethnicity, religion, and ideology, which in turn exacerbate tensions and contribute to social fragmentation.

When considering the utilization of social media platforms by political parties, it can be contended that these parties have extensively employed them to ignite political activism and mobilization in Pakistan, particularly evident during the 2013 general elections. A study conducted by Ali and Gul (2014) shed light on the utilization of Twitter by political parties and activists for campaign-related purposes and real-time updates throughout the elections. The progression of political parties' social media wings since 2013 has been characterized by expansion, technological advancements, integration with traditional campaigning methods, engagement with diverse audiences, and adaptation to regulatory challenges. This evolution underscores the escalating importance of social media in shaping political discourse and mobilization in Pakistan. In recent years, platforms such as Instagram and Snapchat have garnered popularity among Pakistani youth.

Social media use by two major political parties

Since 2013, the development of social media departments within political parties has been characterized by notable expansion and sophistication, underscoring the escalating significance of social media in political discourse and mobilization. Political parties in Pakistan have broadened their digital footprint, establishing dedicated social media divisions to oversee their online endeavors. As outlined by Khan (2016), entities such as the Pakistan Tehreek-e-Insaf (PTI) and the Pakistan Muslim League-Nawaz (PML-N) have made substantial investments in constructing robust social media teams aimed at engaging with supporters and amplifying their messages across platforms such as Facebook, Twitter, and Instagram. The advancement of social media departments has entailed the adoption of cutting-edge digital technologies and resources for online campaigning and communication. As observed by Hussain (2017), political parties

have harnessed data analytics, targeted advertising, and multimedia content production to effectively engage and mobilize voters through social media platforms.

Social media wings now form integral elements of political parties' comprehensive campaign strategies, augmenting conventional campaigning techniques. As per the findings of research conducted by Saeed et al. (2019), parties have melded social media outreach with activities such as rallies, door-to-door canvassing, and traditional media advertising, thereby establishing a cohesive and multi-faceted approach to engaging with voters. These social media wings have evolved to interact with diverse demographics, encompassing urban and rural communities, youth, women, and minority groups. This inclusive approach underscores parties' acknowledgment of the necessity to resonate with a wide spectrum of voters through tailored messaging and outreach initiatives (Ahmad, 2018).

The development of social media wings has additionally been influenced by regulatory hurdles and limitations, encompassing government censorship and constraints on online content. As Butt (2015) elucidates, political parties have adjusted their social media approaches in light of regulatory pressures, resorting to encryption tools and alternative platforms as mechanisms to navigate censorship and guarantee the dissemination of their messages.

When examining the utilization of social media by the Pakistan Tehreek-e-Insaaf (PTI), it becomes evident that PTI significantly expanded its presence across various social media platforms during the specified timeframe. As outlined in a report by Dawn, the party actively interacted with its supporters on platforms such as Facebook, Twitter, and Instagram, leveraging social media channels to mobilize volunteers, disseminate campaign messages, and orchestrate rallies (Dawn, 2014). The leadership of PTI adeptly employed strategic communication and messaging to effectively harness the potential of social media platforms. As highlighted by Hussain and Ali (2017), the party capitalized on social media as a means to communicate its anti-corruption and reformist agenda, resonate with young voters, and challenge conventional political narratives.

PTI seamlessly fused its social media endeavors with conventional campaigning techniques. An article published in The Express Tribune illustrates how the party synchronized online activism with offline initiatives, including door-to-door canvassing and public rallies, to establish a unified and multi-faceted approach to political mobilization (The Express Tribune, 2015). PTI actively interacted with youth and urban populations through tailored social media initiatives. As per a report by Geo News, the party leveraged interactive content, memes, and hashtags to engage young voters and urban demographics effectively, thereby bolstering its electoral appeal (Geo News, 2016).

Harnessing the power of youth and their social media activism, the party's social media mobilization emerged as a pivotal factor in addressing political challenges and controversies. Research conducted by Saleem and Rehman (2015) underscores the party's adept utilization of social media to galvanize supporters, rebut negative propaganda, and garner public backing throughout electoral campaigns.

PML (N) leveraged social media through the establishment of Maryam Nawaz's Media Cell within the Prime Minister House from 2013 to 2016. Under Maryam Nawaz's leadership, the Media Cell substantially expanded PML (N)'s footprint across various social media platforms during her tenure. As reported by Geo News (2016), the cell actively managed numerous social media accounts, including Facebook, Twitter, and Instagram, to interact with supporters and disseminate the party's messages. The Media Cell strategically curated messaging to advance PML (N)'s agenda and counter opposition narratives. An article titled, "Maryam Nawaz sets up media cell in PM House", published by The News International (2013) highlights Maryam Nawaz's direct involvement in overseeing social media campaigns, ensuring alignment with the party's objectives and effectively communicating its policies and accomplishments).

Maryam Nawaz's Media Cell seamlessly integrated social media initiatives with traditional media strategies. As analyzed by The Express Tribune (2014), the cell synchronized messaging across television talk shows, press

releases, and social media platforms to optimize PML (N)'s media presence and influence public opinion. Furthermore, the Media Cell actively engaged with youth and urban audiences through targeted social media campaigns. Mahmood and Malik (2015) elucidate how the cell employed interactive content, memes, and hashtags to capture the attention of young voters and urban demographics, thereby enhancing PML (N)'s electoral appeal. Consequently, it played a pivotal role in addressing political challenges encountered by the PML (N) government. According to a report by Dunya News (2015), the cell adeptly countered opposition narratives and managed public perception during controversies and crises, underscoring its efficacy in shaping the party's image and sustaining public support.

Case studies on digital hate speech against journalists in Pakistan

The preceding overview illustrates how political parties have allocated resources to systematically organize their social media influence to garner support from like-minded individuals. It underscores that anyone, including journalists, media organizations, or even state institutions, can become targets of hate speech by individuals fostering hatred. In a seminal study conducted by Li and colleagues in Pakistan (2023), the researchers examined the patterns of online harassment experienced by political journalists in Pakistan on the Twitter platform, with particular emphasis on the conduct of users affiliated with various political affiliations. By selecting the 12 most active journalists on Twitter, researchers conducted both content and textual analyses to scrutinize the comments posted on their tweets. Employing the theoretical framework of the practice-based theoretical approach, the findings reveal that these journalists predominantly encounter negative comments, including personal attacks, culturally sensitive expressions, and assaults on their professional integrity.

The findings of this study further highlight that commenters aligned with the populist political party, Pakistan Tehreek-e-Insaaf, exhibited a higher tendency to engage in uncivil comments compared to those associated with traditional political parties. Intriguingly, both male and female journalists

received a comparable volume of comments on their tweets; however, the nature of these comments varied significantly. Male journalists were predominantly criticized for their perceived lack of professionalism, whereas female journalists were subjected to offensive comments targeted at their gender. The authors mention that their findings were in line with evidence from other countries like the USA, India, the UK and the Philippines where researchers found that journalists were increasingly harassed on the online platforms.

In another study conducted by Hussain et al., (2022) on the online harassment of female journalists in the backdrop of a petition filed by 128 female journalists in Pakistan on September 7, 2020, alleging that they were subjected to online harassment by political trolls linked with the government. They maintained they experienced gender-based slurs and threats on their timelines, leading to many female journalists leaving Twitter accounts or refraining from engaging online. The results of the study indicate that female journalists were inundated with a significant volume of hateful comments characterized by threatening and offensive language. In addition to facing culturally sensitive remarks and insinuations, the professional integrity of journalists came under scrutiny. Findings of the study largely correspond with research conducted in various regions across the globe, where scholars have identified a pervasive atmosphere of intimidation.

By presenting summarized findings of the above two studies in the context of Pakistan, we intend to examine how journalists become victims of hate speech in Pakistan. The finding provide us a very important aspect that trend of online harassment and hate speech against journalists have emerged as a global phenomenon, especially with regard to populist regimes in various countries. The second important aspect is that since the rise of PTI in the country, media houses started taking sides blatantly which ultimately gave rise to political parallelism in Pakistan, where media outlets align themselves with specific political interests or parties.

The findings from the aforementioned research studies offer insights into the online harassment experienced by journalists during the tenure of the Pakistan Tehreek-e-Insaaf (PTI). These journalists were critical of the PTI government. However, it is pertinent to examine the situation of pro-PTI journalists following the toppling of the PTI government in 2022. Among these journalists from prominent TV channels were Arshad Sharif, Moeed Pirzada, Imran Riaz Khan, and several others who transformed into vehement critics of the newly formed PDM government and the military establishment. In response, they encountered backlash from the newly established government. Arshad Sharif tragically met his demise in Kenya, while Imran Riaz Khan faced arrest by law enforcement agencies on two occasions, and Moeed Pirzada fled to the U.S. Interestingly, Imran Riaz Khan admitted in his Vlog that he received all corruption-related information from intelligence operatives and disseminated the information they provided. Similarly, Moeed Pirzada confessed during an interview with renowned journalist and YouTuber Mattiullah Jan (MJTV on YouTube) that he received information from intelligence officials and considered them as trusted sources.

The results gleaned from the aforementioned two research studies, which span the tenure of the PTI government, along with the accounts concerning pro-PTI journalists during the period of the PDM government, point to significant concerns regarding media and journalistic ethics within the country. These practices underscore that neither the government, political parties, nor media entities and journalists are operating in the best interests of the public and democracy.

In addition to the aforementioned considerations, it is imperative to reintegrate the masses (public/audience) into the broader discourse surrounding political communication research, a facet that has somewhat receded in this discussion. Undoubtedly, democracy, government, and journalists exist to serve the public, and as the fourth estate, the media and journalists bear the paramount responsibility of acting in the public interest rather

than aligning with factions and adopting propagandist stances. They must uphold the principles of objectivity and impartiality. Therefore, it is essential to revisit foundational principles, particularly as the current social media landscape has empowered the audience/public with a feedback mechanism to scrutinize and challenge the statements made by government officials, politicians, and even journalists.

Social media platforms offer instantaneous feedback mechanisms, enabling audiences to respond to and influence news stories as they unfold. Journalists and media organizations can monitor audience reactions in real-time and adapt their coverage accordingly. This immediate feedback loop furnishes audiences with a potent tool for shaping media narratives. A study conducted by Tandoc et al. (2017) delves into the concept of "constructive journalism", wherein journalists actively engage with audience feedback on social media to enhance the quality and relevance of their reporting.

Given the widespread issue of online harassment targeting journalists on a global scale, we examine the literature concerning the influence of journalists' political affiliations on their professional conduct in other regions. In Western democracies, recent scholarly works have shed light on the effects of journalists' political leanings within the framework of progressively polarized media landscapes. For instance, research by Gentzkow and Shapiro (2011) and Prior (2013) has explored the phenomenon of partisan media platforms catering to particular ideological demographics, resulting in echo chambers wherein individuals encounter information that reinforces their existing beliefs.

Moreover, investigations conducted by Stroud (2011) and Iyengar and Hahn (2009) have delved into the concept of selective exposure, wherein individuals actively seek news sources that mirror their political inclinations. This pattern carries substantial ramifications for journalists, as they might encounter pressure to adhere to the editorial biases of their affiliated outlets to maintain their audience base and financial sustainability.

Regarding Asian democracies, contemporary literature has centered on the obstacles confronting journalists functioning within contexts marked by restricted media freedom and governmental censorship. For instance, research conducted by Repucci (2020) has chronicled the deterioration of press liberty in nations like China, Russia, and Turkey, where journalists face the looming threats of incarceration, intimidation, and even physical assault due to their journalistic endeavors.

Additionally, investigations conducted Hanitzsch et al. (2020) have underscored the widespread existence of state-controlled media and the stifling of independent journalism across numerous Asian democracies. This situation has raised apprehensions regarding the caliber and precision of news reporting, as journalists may engage in self-censorship or adhere to governmental directives to evade retaliatory actions.

In both Western and Asian democracies, the emergence of digital media and social networking platforms has added layers of complexity to the journalistic environment. Investigations conducted by Tsfati et al. (2014) and Newman et al. (2020) have delved into the repercussions of social media on patterns of news consumption and the dissemination of misinformation, thus affecting journalists' gatekeeping responsibilities and adherence to professional standards.

In general, while the influence of journalists' political affiliations is undeniable in both Western and Asian democracies, the distinct dynamics and obstacles differ significantly between the two regions. Western democracies often grapple with the ramifications of partisan polarization and audience segmentation, while Asian democracies face significant challenges related to media suppression and censorship.

Conclusion

In Pakistan, media organizations frequently associate themselves with particular political agendas, resulting in polarization and discord within the media sphere (Khan, 2019). This alignment fosters a climate where

journalists perceived to have ties to specific political factions become subjected to hate speech and online harassment (Ahmad, 2017). Social media platforms have been exploited as tools to disseminate hate speech targeting journalists who are perceived to be critical of particular political parties or narratives (Gul, 2018). This trend is exacerbated by the widespread dissemination of fake news and misinformation campaigns, which intensify tensions and fuel animosity towards journalists (Ali, 2020).

The widespread occurrence of hate speech targeting journalists on social media platforms has a stifling impact on freedom of expression and press liberty in Pakistan (Raza, 2016). Journalists encounter intimidation, threats, and online assaults, resulting in self-censorship and hesitancy to cover contentious topics (Zaidi, 2021; Hussain, 2022). Despite legal safeguards for journalists, enforcement mechanisms are inadequately enforced, allowing those responsible for social media hate speech to operate with impunity (Butt, 2019). The ineffective legal redress exacerbates journalists' susceptibility to online abuse and harassment (Malik, 2018).

Political figures and influential individuals wield considerable influence in fanning the flames of social media hate speech against journalists through their rhetoric and public pronouncements (Khalid, 2020). Their provocative comments foster an environment of animosity towards journalists and diminish the media's watchdog function in society (Hussain, 2018).

In general, the prevalence of political parallelism in Pakistan has fostered a hostile atmosphere where journalists are increasingly subjected to social media hate speech. Tackling this challenge necessitates collaborative initiatives from the journalistic community, governmental bodies, and civil society to safeguard press freedom and guarantee journalists' safety in the digital realm. The Pakistani government has initiated measures to oversee social media platforms and tackle concerns regarding hate speech and misinformation. In 2020, the government implemented the Citizens Protection (Against Online Harm) Rules, designed to govern social media platforms and address issues related to online abuse and hate speech (Government of Pakistan, 2020).

In summary, social media serves as a pivotal platform for the propagation and normalization of hate speech in Pakistan. The platform's functionalities, algorithmic biases, and difficulties in content moderation contribute to the spread of divisive narratives and extremist viewpoints. Combating hate speech on social media necessitates a collaborative approach involving government regulation, advocacy by civil society, and responsible governance by platforms to uphold digital rights, foster tolerance, and promote diversity.

The chapter underscores the intricate relationship between social media usage, journalists divergent from ethical norms and political polarization in Pakistani society. While social media platforms offer avenues for political expression and participation, they also present significant challenges in terms of fostering echo chambers, disseminating misinformation, and amplifying divisive rhetoric. Addressing these challenges demands joint efforts from journalists associations, policymakers, civil society groups, and technology firms to enhance digital literacy, counter misinformation, and facilitate constructive dialogue across political divides.

References

Ahmad, S. (2017). Media and politics in Pakistan: A case study of political parallelism. *Journal of Media Studies, 14*(2), 78-93.

Ahmad, T. (2018). Political communication in Pakistan: A case study of general elections 2018. Pakistan Journal of Commerce and Social Sciences, 12(3), 896-915.

Ali, A. & Gul, S. (2014). Social media: A platform for political activism in Pakistan. *International Journal of Asian Social Science, 4*(10), 961-970.

Ali, M. (2020). Social media hate speech and its impact on journalists in Pakistan. *International Journal of Communication, 15*, 205-220.

Butt, A. (2019). Legal challenges in combating social media hate speech against journalists. *Pakistan Law Review, 5*(1), 45-60.

Butt, S. A. (2015). Social media and political marketing in Pakistan: Opportunities and challenges. South Asian Studies, 30(2), 353-368.

Dawn. (2014, May 18). PTI's social media strategy pays off. *Dawn.* https://www.dawn.com/news/1107173

Dawn. (2020, August 13). *Trolls and hate. Dawn.* https://www.dawn.com/news/1574124

Digital Rights Foundation. (2013). *Freedom on the Net 2013: A global assessment of internet and digital media* (S. Kelly, M. Truong, M. Earp, L. Reed, A. Shahbaz, & A. Greco-Stoner, Eds.). Digital Rights Foundation. https://freedomhouse.org/sites/default/files/2020-02/FOTN%202013_Full%20Report.pdf

Dunya News. (2015, August 8). Maryam Nawaz's media team responds to opposition allegations. *Dunya News.* https://dunyanews.tv/en/Pakistan/292232-Maryam-Nawaz-media-team-responds-to-opposition-allegations

Gentzkow, M. & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics, 126*(4), 1799-1839.

Geo News. (2016, August 29). Maryam Nawaz media cell revamps PML-N's social media. *Geo News.* https://www.geo.tv/latest/112462-Maryam-Nawaz-Media-Cell-revamps-PML-Ns-social-media

Geo News. (2016, August 29). *PTI's social media team woos youth.* https://www.geo.tv/latest/112462-ptis-social-media-team-woos-youth

Gul, F. (2018). Weaponization of social media: A threat to journalistic freedom in Pakistan. *Journal of Social Media Studies, 7*(2), 112-128.

Hanitzsch, T., Hanusch, F., Ramaprasad, J., Anikina, M., Berganza, R., Cangoz, I., Coman, M., Hamada, B., Hernandez, N., Karadjov, C. D., Moreira, S. V., Mwesige, P., Plaisance, P. L., Reich, Z., Seethaler, J., Skewes, E. A., & Wang, Q. (2020). Worlds of journalism: Journalistic cultures, professional autonomy, and perceived influences across 18 Nations. *Journalism Studies, 21*(4), 461-481.

Human Rights Watch World. (2021). *Pakistan: Escalating attacks on journalists.* https://www.hrw.org/news/2021/06/03/pakistan-escalating-attacks-journalists

Human Rights Watch. (2022). *Pakistan: Events of 2022.* https://www.hrw.org/world-report/2023/country-chapters/Pakistan

Hussain, S. & Ali, S. (2020). Social media, polarization and political participation: A case study of Pakistan. *Pakistan Journal of Social Sciences*, *40*(2), 555-572.

Hussain, S. (2017). The role of social media in the electoral politics of Pakistan. Pakistan Journal of Commerce and Social Sciences, 11(1), 99-118.

Hussain, S. (2022). Analyzing media–government relations on policy issues in the semi-democratic milieu of Pakistan. *Journalism*, *23*(8), 1770-1788. https://doi.org/10.1177/1464884920969086

Hussain, S., Bostan, H., & Qaisarani, I. (2022). Trolling of female journalists on Twitter in Pakistan: An analysis. *Media International Australia*, *191*(1), 129-146. https://doi.org/10.1177/1329878X221145977

Iyengar, S. & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*(1), 19-39.

Khalid, A. (2020). The role of political leaders in inciting hate speech against journalists. *Journal of Political Communication, 25*(4), 321-336.

Khan, A. N. (2016). Role of social media in Pakistani politics: Trends, benefits and risks. *South Asian Studies*, 31(1), 261-274.

Khan, Z. (2019). Political polarization and its impact on media freedom in Pakistan. *Pakistan Journal of Communication Studies, 11*(3), 176-192.

Li, M., Hussain, S., Barkat, S., & Bostan, H. (2023). Online harassment and trolling of political journalists in Pakistan. *Journalism Practice*, 1-18. https://doi.org/10.1080/17512786.2023.2259381

Mahmood, S. & Malik, A. (2015). Social media for political marketing in Pakistan: A case study of Pakistan Muslim League (Nawaz) social media campaign for general elections 2013. *Journal of Political Studies*, 22(2), 93-112.

Malik, K. (2018). Legal protections for journalists in Pakistan: A critical analysis. *Pakistan Law Journal, 10*(2), 89-104.

Ministry of Information Technology and Communication. (2020). *Citizens protection (against online harm) rules, 2020*. Government of Pakistan. https://digitalrightsfoundation.pk/wp-content/uploads/2020/02/CP-Against-Online-Harm-Rules-2019-updated-1.pdf

Mughal, N. (2022, September 19). In Pakistan, legal action, online threats leveled at political reporter. *Voice of America*. https://www.voanews.com/a/in-pakistan-legal-action-online-threats-leveled-at-political-reporter-/6754261.html

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2020). *Reuters Institute Digital News Report 2020*. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf

Pakistan Institute of Legislative Development and Transparency. (2019). *Democracy and the role of political parties in Pakistan*. https://www.pildat.org/wp-content/uploads/2019/08/Democracy-and-the-Role-of-Political-Parties-in-Pakistan-English.pdf

Prior, M. (2013). Media and political polarization. *Annual Review of Political Science*, *16*, 101-127.

Raza, A. (2016). Social media hate speech and freedom of expression in Pakistan. *Journal of Media Ethics, 8*(1), 55-68.

Reporters Without Borders (2021). *Online harassment of journalists*. rsf_report_on_online_harassment.pdf

Repucci, S. (2022). *Freedom in the World 2020: A leaderless struggle for democracy*. Freedom House. https://freedomhouse.org/sites/default/files/2020- 02/FIW_2020_REPORT_BOOKLET_Final.pdf

Saeed, A., Shaukat, N., & Azhar, M. (2019). Role of social media in political mobilization: A case study of Pakistan. *Journal of Political Studies*, *26*(1), 121-138.

Saleem, H. & Rehman, S. (2015). Social media and political participation: A case study of Pakistan. *Journal of Political Studies, 22*(2), 81-98.

Saleem, H. & Rehman, S. (2017). Role of social media in political polarization: A case study of Pakistan. *Pakistan Journal of Commerce and Social Sciences*, *11*(3), 882-895.

Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press.

Tandoc, E. C., Ferrucci, P., & Duffy, M. (2017). Facebook use, news sharing, and perceptions of information overload: A survey of U.S. users. *Digital Journalism*, *5*(10), 1344-1363. https://doi.org/10.1080/2167081 1.2017.1298494

The Express Tribune. (2014, September 14). *Maryam Nawaz reshapes PML-N's media strategy*. https://tribune.com.pk/story/756291/maryam-nawaz-reshapes-pml-ns-media-strategy/

The Express Tribune. (2015, May 5). *PTI's social media power*. https://tribune.com.pk/story/879582/ptis-social-media-power/

The News International. (2013, October 20). *Maryam Nawaz sets up media cell in PM House*. https://www.thenews.com.pk/print/63544-maryam-nawaz-sets-up-media-cell-in-pm-house

Tsfati, Y., Stroud, N. J., & Chotiner, A. (2014). The impact of media skepticism on political participation. *International Journal of Public Opinion Research*, *26*(3), 275-299.

Tufail, A., Nawaz, H., & Rana, A. (2018). Role of social media in political polarization: A case study of Pakistan. *Journal of Political Studies*, *25*(1), 225-244.

Zaidi, F. (2021). Self-censorship among journalists in the face of social media hate speech. *Journal of Journalism Studies, 14*(4), 287-302.

FACT-CHECKING ON TWITTER IN THE FACE OF FAKE NEWS: COMBATING HATE SPEECH ON SOCIAL NETWORKS

Enrique Javier Díez-Gutiérrez
/ University of León, Spain

María Verdeja
/ University of Oviedo, Spain

Eva Palomo-Cermeño
/ Rey Juan Carlos University, Spain

Introduction

It seems increasingly difficult to discern between what is true and what is false in the face of so much noise on social networks. Disinformation has become a real problem in the 21st century (Guallar et al., 2020).

The term fake news usually refers to false news or hoaxes that are spread through social networks and media, but it could also be defined as a type of messages that, without having to be completely false or untruthful, are versions or interested stories that lack rigor, credibility and veracity in their contents; their purpose is to be disinformation affecting our perception of reality (Alonso, 2019; Losada, 2020). More than a specific type of content, disinformation is a process where fake news is mixed with truthful information and where different channels cooperate to enhance the credibility of such malicious accounts (Elías, 2021).

In this regard, the European Commission recalls that it is informative content "false, inaccurate or misleading

[...] intentionally designed, presented and promoted to cause public harm or private gain" (European Commission, 2018: 10).

This phenomenon is causing a worrying 'information epidemic' (Salaverría et al., 2020), especially when research confirms that 86% of the Spanish population is unable to identify and recognize fake news (Amorós, 2018) and 68% admit they are concerned about this issue (Reuters Institute, 2019).

This 'infodemic' (Popiolek, Hapek, & Barańska, 2021) has turned social networks into one of the most widely used channels of spreading hoaxes (Casero-Ripollés, 2020; Guallar et al., 2020) exponentially increasing the problem of disinformation, the consequences of which have even been warned about by the European Commission (Nieto, 2021). Moreover, it must also be taken into account that hoaxes and rumours tend to generate more attention and diffusion than truthful news (Vosoughi, Roy, & Aral, 2018).

Advertising techniques applied to social networks play a very important role in the virality of the spread of fake news and the extension of digital noise. One of the originators of this method is Stephen Kevin "Steve" Bannon. He was a publicist. Advertising alludes precisely to this type of emotional stimuli (not rational) with many allusions to the sexual theme, power, image, etc. Cambridge Analytica is a parent company of another much larger one called SCL Group, which collaborates with the Pentagon, the United States military, digital platforms, etc. Following Steve Bannon's approach, these groups have been studying for more than 20 years how this type of digital stimuli changes the behaviour of large masses of population.

One of the most extensive studies on news dissemination conducted by the Massachusetts Institute of Technology (MIT) and the Sloan School of Management, which analysed 126,000 news items on social networks from 2006 to 2017, found that fake news is retweeted 70% more on average than real news (Vosoughi, Roy, & Aral, 2018); mainly because it seems fresher, more striking in the way it is presented, and above all it appeals to emotion which helps it to spread and go viral faster.

The purpose of this research was to analyse the fact-checking work on Twitter conducted by a professional from the Pandemia Digital account. Specifically, the analysis was carried out in relation to the controversy surrounding macro-farms, which went viral with the hashtag #MacrogranjasNO based on the declaration of the Spanish Minister for Consumer Affairs defending extensive livestock farming.[1]

Fact-checking

Faced with this pressing problem of online noise which prevents access to the right to truthful public information established in article 19 of the Universal Declaration of Human Rights and recognized in the Spanish Constitution, the so-called "fact-checking" has emerged with the intention of checking the information published in the media and social networks in order to detect errors and false news and determine its veracity and correctness, making errors, lies and inaccuracies visible (Ufarte-Ruiz, Peralta-García, & Murcia-Verdú, 2018).

Research on fact-checking activity has been conducted from different approaches and perspectives: epistemological analysis (Graves, 2018), models (Singer, 2018), procedures (Lim, 2018), impact and effectiveness (Young et al., 2018), transparency (Humprecht, 2019), presentation of results (Vizoso & Vázquez-Herrero, 2019) or verification formats (Ecker et al., 2019; García--Crespo, Ramahí-García, & Dafonte-Gómez, 2021).

One of the most important verification initiatives worldwide is the International Fact Checking Network (IFCN). This initiative was launched in 2015 and is led by the Poynter Institute for Journalistic Studies. Subsequently, specialised groups have emerged that carry out this work in an organized manner; many of them are grouped in the International Fact-Checking Network which supports verification platforms worldwide.

Its purpose is: "to bring together the main verification agencies arising in the different countries of the world and to promote good practices".

A study has recently been published (Carvajal et al., 2022) that places fact-checking among the 20 innovations with the greatest industrial and social impact of the last decade in Spain, but we must say that in Spain fact checking is still in an incipient phase (López-Pan & Rodríguez, 2020).

There are currently three news verification agencies that are members of the IFCN: Newtral, EFE Verifica and Maldita.es. Entering the verification bodies and following the classification made by Graves (2016), they could be classified into three categories: a) promoted by civil society, b) linked to the media, and c) autonomous journalistic websites.

Fact-checking on Twitter

Given that fake news spreads much faster on social networks than in traditional social media (Nieto, 2021), in this research we analyse fact-checking on Twitter, since it is one of the social networks with the greatest capacity to "viralise" unverified hoaxes (Córdoba-Cabús & López-Martín, 2021; Gómez--Calderón, 2020), even through trolls and bots[2] (Jamison, Broniatowski, & Quinn 2019) with the purpose of manipulating and generating disinformation and confusion among the population (Sharevski, Jachim, & Florek, 2020), but where disinformation is also being fought more (López-Martín & Córdoba-Cabús, 2021).

Although Twitter presents a relatively moderate volume of use, its debates and controversies sometimes provoke highly stimulated emotions. These are more likely to be shared given that the controversy is overrepresented and the traditional media reproduce them and some even end up setting

2. In the Internet jargon, *troll* refers to participants with unknown identity who post provocative messages in a discussion in an online community, with the intention of annoying or provoking a negative emotional response in the other participants, trying to make them angry and confront each other. A *bot* is a computer program that automatically performs repetitive tasks on the Internet and is used in social networks to simulate human interaction artificially inflating the number of visits or followers, or automating responses to position messages or influence debates.

the social, media, and political agenda by establishing "issues, trends, and positions in the public debate" (Hernández-Conde & Fernández-García, 2019: 36).

Obtaining information from social networks is the most common use by people between 16 and 64 years old (Digital, 2021). Young people in Spain predominantly choose social networks as sources of information (Reuters Institute, 2020), but in addition, 70% of journalists admit using them as a source for writing news, Twitter (87.90%) being the most used channel (Janssen-Observer, 2017). Twitter stands out as the social network with a large percentage of users interested in current issues. This network has become "a viral medium that facilitates the rapid circulation and multiplication of messages" (Paz-García & Spinosa, 2014: 136). It can also be said that it is one of the social networks where dialectical polarization has grown the most in recent years as well as hate messages and where lies have increased in digital conversations (Domínguez, 2021).

Methodology

The methodology employed is qualitative and the main technique used for collecting information was an in-depth interview (Castro, 2021) with Julián Macías-Tovar, head of the Digital Pandemic platform. He is a professional fact-checker who has a verification platform from which he dismantles informative hoaxes circulating on social networks.

To complete this information, we also relied on messages – tweets – published by the verification platform of the fact-checker interviewed, all with the aim of understanding how fake news works on Twitter. The Twitter account Digital Pandemic was chosen because it is a fact-checking platform specialised in combating hoaxes and fake news of a political nature that have had great repercussions recently. In Table 1 we collect the questions asked during the interview.

Table 1. Interview questions

Dimensions	Questions
Impact of fake news	How does fake news work on Twitter and how does it go viral?
	Does it really generate noise on the network or only in like-minded groups?
	Is fake news disinformation, manipulation or does it also entail hatred and confrontation?
Fact-checking	Why does fake news "hook" more than real news?
	Is it possible to avoid them or, at least stop them with fact-checking?
	Does fact-checking that questions an ideology reach those who are sympathetic to it?
Political use in democracy	What kind of political use is made of fake news and what are its aims?
	In the case of macro-farms, what was intended and what has been achieved?
	Does fake news challenge democracy or endanger it?
	What political, collective, and public measures and alternatives should be taken in a democracy in Spain, in the European Union and worldwide?

Source: Elaborated by the authors

The case of the #MacrogranjasNO campaign

One of the fake news that has had the greatest impact in Spain, and even in Europe, has been around macro-farms; i.e. intensive livestock farming facilities that host thousands of head of cattle fed with animal feed and are stabled in industrial warehouses.

On 26 December 2021, the English newspaper The Guardian published an interview with the Spanish Minister of Consumer Affairs, Alberto Garzón, on livestock farming in Spain. In it he stated that compared to ecologically sustainable extensive livestock farming, intensive livestock farming and macro-farms meant "taking a village in depopulated Spain, putting 4,000 head of cattle there – or 5,000 or 10,000 –, contaminating the soil, polluting

the water and then normally exporting it ... It is poorer quality meat, it is animal abuse and it has a huge and disproportionate ecological impact."

His statements had gone unnoticed until 3 January 2022 when a meat industry lobby website launched a campaign against the minister assuring that "Garzón affirms in The Guardian that Spain exports poor quality meat from mistreated animals." It was not true, but it didn't matter.

The European Commissioner for Agriculture, Janusz Wojciechowski, intervened by reminding the public that large-scale farms were a "problem" in Europe and pointed out in the midst of the controversy that the EU wants to promote small-scale extensive livestock farming.

Despite this, the Popular Party (Partido Popular, PP) decided to use this hoax in motions before institutions of all Spain demanding the resignation of Minister Garzón. Simultaneously, this was used in their electoral campaign.

The last consequence that we have reviewed in this investigation is the violent assault on the Town Hall of Lorca in the Region of Murcia. Dozens of people breaking the police cordon and shouting "we are going to kill you" managed to paralyze and cancel the municipal plenary session where a municipal regulation was to be ratified to prohibit the construction of new macro-farms near urban centres, schools and springs. The regulation was the ratification of an agreement that had already been unanimously approved by the Lorca City Council itself in 2020 – PP, PSOE, IU, Greens, Ciudadanos and VOX – after months of negotiations with the livestock farming sector and neighbourhood platforms (Calvo & Reche, 2022).

The mayor of Lorca described this action as "an attack on democracy." The parliamentary group Unidas Podemos called it "21st century fascism" and has linked it to the atmosphere of "hoaxes" and "intoxication" spread by "right-wing and extreme right-wing parties". The PP expressed its general condemnation of "all kinds of violence." On the other hand, the far-right party VOX expressed its support (Otero, 2022) and managed to get the Regional Parliament of Murcia not to condemn the assault (Calvo & Reche, 2022).

Pedro Giner, a VOX militant and one of the protagonists of this action, claimed that they were victims of disinformation. "They told us things that did not coincide with reality and the atmosphere became very heated. Disinformation often makes us take decisions that are not coherent, and this is one of them" (Almagro & Cabrera, 2022).

Results of the in-depth interview

We present the results obtained from the in-depth interview with Julián Macías-Tovar.

Viral fake news and digital noise

One of the substantial problems of fake news is its virality. In the case of macro-farms, it was indeed very fast and highly effective. According to Macías-Tovar the three factors involved are the following:

> Regarding how fake news works on Twitter and how it goes viral, it seems that it has to do with the combination of three key elements: 1) the attraction of lies, 2) the link with strong feelings and emotions such as hatred, and 3) viralisation with the help of fake accounts or bots. (Interview)

Figure 1. Elements of viralisation of fake news
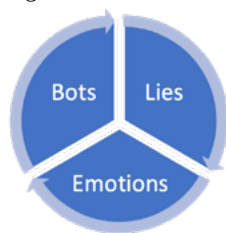
Source: Elaborated by the authors

As Macías-Tovar points out, the lie itself is not so much what causes viralisation, but rather its connection with emotional reactions linked to surprise, indignation, hatred, etc.:

An example of this can be seen in the Alberto Garzón issue with headlines such as "Spain exports poor quality meat and abuses animals". This headline angers people – it does not matter if it is true or false –. (Interview)

Figure 2. Fake news headline

Garzón asegura que España exporta carne «de mala calidad de animales maltratados»

El ministro de Consumo ataca las macrogranjas que «contaminan el suelo y el agua»

El ministro de Consumo, Alberto Garzón. / r. c.

Source: El Comercio (Asturias): https://cutt.ly/6PbK4ef

Translation: Garzón claims Spain is exporting meat "of low quality from mistreated animals": The minister of Consumer Affairs attacks macro-farms that "pollute the soil and the water".

The key, Macías-Tovar insists, is an emotional issue. What is interesting is the wrapping, the communicative framework in which the lies are packaged. If there is an emotional component such as "hate" viralisation has a much greater effect:

The strategy is to generate hatred, to viralise communicative frames. What is really effective is to make people angry. (Interview)

In addition to these two issues – lies and emotions such as hate – there is a third key factor: virality, and it is possible thanks to what Macías-Tovar points out: the funding behind many of these campaigns.

> A large part of the funding of these campaigns goes to bots that spread the message automatically and instantaneously through the networks in order to give an image of a massive and spontaneous collective event. (Interview)

A characteristic element of detection of automated bots in these campaigns is the capacity for simultaneous replication of the same message with the same typo or error by thousands of accounts at the same time, given that it is materially impossible for more than 30,000 people to make the same mistake at the same time (Calvo, Campos-Domínguez, & Díez-Garrido, 2019). These bots are often aided and abetted by the massive creation of fake accounts.

As Macías-Tovar explains, an element of detection of these accounts is that they tend to be accounts with eight digits, as when many accounts are created, Twitter automatically grants by default a user ending in eight digits, which requires human intervention to change it later. It is also the case of new accounts recently created, i.e. those with less than a year of existence and which are used especially for this type of campaigns (Luque, Macías-Tovar, & Casado, 2021).

## Fake news beyond the networks: Political use and harassment of democracy

Digital campaigns alone have very little impact. They have the capacity to reach small populations and make people believe fake news, but they cannot reach large masses. The big impact comes when the media and civil society come into play:

> The most important political element, I think, in any country, is the media, in coexistence with the digital media – which in a way are also the media –. They are the entity with the capacity to generate moods in the population; moods that can turn into a coup d'état. They can also mobilise votes, etc. (Interview)

A simultaneous effect of the viralisation of these campaigns on social networks is their projection and repercussion in traditional media:

> If the campaign against Garzón had only been on social networks, surely it would have been very small, but the disinformation campaign has a triple strategy: 1) the digital one, shared in networks through impact headlines, bots, etc. 2) the media, replicated by like-minded media (press, television), and 3) the real one, that of civil society. (Interview)

Figure 3. Disinformation strategy
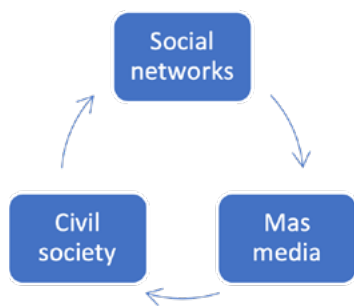
Source: Elaborated by the authors

In this way, we can observe how these hoaxes end up going beyond social networks and into real life, which is really the aim of this type of campaign, i.e. the objective is not only to promote digital noise, but to provoke conflict in the real world:

> In relation to what happened in the Lorca Town Hall when they arrested one of the leaders of the assault, he claimed that he was very much uninformed because he had been told it was going to harm the entire livestock sector. (Interview)

In this case, the intervention of political opposition groups played a prominent role, launching a campaign on Twitter with the hashtag #MásganaderíaMenoscomunismo (#MoreLivestockLessCommunism), on the occasion of the elections in the Autonomous Community of Castilla y

León. It was meant to question the role of the Minister of Consumer Affairs and to conveniently call for his resignation.

In the digital and information age, these "battles" in social networks and media, in coordination with political actions and those of foundations or social groups have been called digital hybrid wars (Polanco, 2021).

> The issue of digital hybrid warfare in networks seems to make sense. It is never only digital but also implemented with other strategies, such as those of the media, political actions, etc. All of this is what can be called hybrid warfare and it can be effective when it is well articulated. (Interview)

Fake news, democratic deliberation on the net and democracy

According to Macías-Tovar these types of campaigns question democracy and endanger it. He argues that one of the main problems has to do with the fact that today's media is not transparent:

> The fact that they are increasingly using lies and generating hatred instead of following the deontological code as an element of journalism and telling the truth in a pluralistic way is a problem, because they have the ability to influence society. My conclusion after a long time is that the MOB (Mentira, Odio y Bots) method – Lie, Hate and Bots – is very effective, but in addition there is a fourth element that allows it: total impunity. (Interview)

In this sense, Macías-Tovar explains that even though the networks favour freedom of expression, it seems that they violate other types of rights such as the right to receive truthful information. Furthermore, the lack of political will to fight disinformation is an added factor that significantly increases the problem:

> There is no political will whatsoever. Absolutely nothing is being done, to the point of shame; neither the platforms nor the institutions. (Interview)

Analysing the Digital Pandemic Twitter threads we can see that there were different phases. First there was a phase of harassment and demolition

of Minister Garzón demanding his resignation. As described by Macías-
-Tovar this is a time of confusion in which there are different actors in
different media.

> First, there was a clear attack. There were threads on Twitter in different
> sites that were widely spread, articles in the digital press, etc. (Interview)

In this phase there is a moment of a lot of (dis)information sometimes com-
ing to resemble a "media circus" in which the right and the far right make
all kinds of accusations that have nothing to do with macro-farms:

Figure 4. Tweets by Cuca Gamarra and Iván Espinosa de los Monteros
(Spokespersons of the PP and VOX parliamentary groups in Congress)

Cuca Gamarra ✔
@cucagamarra

Pasadas las 24 horas, ni Sánchez le ha cesado ni
Garzón dimite. En apoyo al sector ganadero,
@GPPopular hemos registrado en el Congreso una
Proposición no de Ley para la reprobación del
ministro al que nadie respalda, la retirada de la
campaña y el apoyo al sector.

Source: @cucagamarra

Iván Espinosa de los Monteros ✔
@ivanedlm

Claro que sí, los comunistas también tienen
derecho a cenar, y a cenar jamón del bueno. Ese
no es el problema.

El problema es que ellos cenen jamón y pretendan
prohibírselo a los demás. Lo de siempre de la
izquierda!

Source: @ivanedlm

Translation: *Tweet by Cuca Gamarra*
*24 hours after neither Garzón has resigned nor has Sánchez (President) made him quit. In su-
pport of the livestock farming sector the @GPPopular has registered a bill in the Congress to
condemn the minister who nobody supports, to demand the withdrawal of the campaign and
support the sector.*

Translation: *Tweet by Espinosa de los Monteros*
*Of course, communists also have the right to dinner and to eat good ham. That's not the pro-
blem. The problem is when they dine with ham while they prevent others from doing so. It's
always the same with the left!*

However, sometimes democracy is defended and when you want to dismantle a lie it is possible to do so. The truth can also be very powerful; for this, the collaboration of the media is essential (Macías-Tovar, 2022).

It is in the second moment where, as a response to everything that was happening, the media began to act through different channels (TV programs, digital press, etc.) they managed to turn the initial situation around:

> Nobody really wants macro-farms (not even the PP). As soon as the media starts talking about macro-farms the truth becomes clear; no one defends them. The key is the participation of the media. (Interview)

Figure 5. Twitter thread by Julián Macías-Tovar

Source: @JulianMaciasT

Translation: *Tweet by Macías-Tovar*
One of the keys to turning the tide was the reaction of dozens of journalists who supported Garzón's position, as well as the incongruence of the rest of the politicians who criticized him, but specially the immovable position of Garzón.

According to Macías-Tovar, one of the problems with fake news on Twitter and the media that contribute to its dissemination is that they are orchestrated media campaigns that aim to get people into the game of denying, refuting, reacting, or angering. This is one of the primary objectives since it allows many more people to join in from the confrontation, from the "other side":

> Twitter and most social networks through their algorithms tend to create ideological bubbles of affinity. The aim of digital noise is not only to

Fact-checking on Twitter in the face of fake news:
Combating hate speech on social networks

create "affinity communities", but also to involve other different ideological bubbles that deny, criticize or directly confront fake news or hate messages and to turn these messages into a trending topic. (Interview)

What matters is the packaging, the communicative framework in which the lie they want people to talk about is wrapped; and if there is a strong emotional component such as "hatred" viralisation has much more effect:

The strategy is to generate hatred, to viralise normative frameworks, communicative frameworks, etc. Therefore, the truth cannot be an impediment for this to happen, but what is really effective is to anger people. (Interview)

Finally, there is a third moment which Macías-Tovar calls "the ridiculous" in which once the truth has been unravelled, there are yet political groups that still want to sow confusion. The issue is now in its final throes:

In a meeting with journalists Pablo Casado is ridiculed when he is asked: Why have you brought us to an extensive cattle ranch to ask for Garzón's resignation, when that is what the minister is really defending? (Interview)

Figure 6. Twitter thread by Julián Macías-Tovar

Source: @JulianMaciasT

Disinformation and digital noise are not only the messages that are transmitted and the manipulation they entail, but also the concealment of certain aspects. For example, at the beginning no politician – PP, VOX, Ciudadanos or PSOE – mentioned the term macro-farms as it is the word they try to hide while they talk about Garzón attacking the entire livestock farming sector.

All of which on this occasion generated the opposite effect. Thus, if at the beginning during the two weeks of the campaign, the politicians did not mention the word macro-farms, once the analysis focused on the essential issue, a very important problem emerged and became the centre of the debate:

> This was turned around and a very striking fact is that after this the word macro-farms was used thousands of times. In fact, during the two weeks the controversy lasted 80% of the mentions in the last 10 years were made. (Interview)

Fighting fake news

Macías-Tovar's analysis is based on the fact that every company has an interest. In other words, no one creates a business to lose money, which is rational in a capitalist economy governed by profit as the driving force of the economic system (Briceño-Montilla, 2022). If a news story goes viral even if it is false, it means more profits for the companies that advertise it. Algorithms benefit hatred in the networks, as reflected in the reports of some former Facebook workers who claimed that the algorithm gave the angry reaction five points more than other reactions – for example, the like which is given one point:

> Twitter, Facebook, YouTube, etc., have an economic component; they are there for their interests but for me they also have a geopolitical

> component, so it is impossible that they will change… but, this does not mean that alternatives cannot be proposed. (Interview)

According to Macías-Tovar, the first proposal at a global level involves greater public control over these technological giants and greater transparency and control over the algorithms behind these social networks:

> I would propose the following measures: first of all, to pay more taxes. These are the richest companies but they are the ones that pay the least taxes. Another issue is transparency on the subject of the algorithm. (Interview)

But the crucial issue is to articulate mechanisms to combat disinformation. To this end he proposes the persecution of those self-styled media that systematically contribute to the dissemination of false news.

> There are media that only live on fake news and how it is spread on social networks, with news that has been shared more than 500,000 times. This means it has been seen by millions of people. When there are media that are only dedicated to lying, you have to close them. (Interview)

Another measure Macías-Tovar proposes is to strengthen current legislation and reach a globally agreed regulatory system to combat impunity of fake news. One of the greatest dangers is that those who spread lies have total impunity unless it can be demonstrated that the fake news they transmit falls under any of the definitions of hate crimes in current legislation (Ministry of the Home Affairs, 2020).

The analysis of hate speech deserves special attention because of its social, cultural and educational implications, its influence on the construction of a social and political climate, and also because of the link between the rise of online hatred and the perpetration of hate crimes (Müller & Schwarz, 2021).

As Macías-Tovar points out, legislation cannot be an accomplice to the impunity with which certain media disseminate false news:

> Consensual Legislation would be necessary to fight against this impunity. It would also be good if there were more pluralistic networks of fact-checkers. (Interview)

In this regard he is very critical of some fact-checkers. He considers that the work done by certain fact-checkers could be very much improved:

> When they detect false news and after it has gone viral they display a little sign saying "this news is false" that is no longer of any use. What I would do is to notify 100% of the people who saw a fake news story that the source has lied. In that way, all the people who have consulted the hoax can see it. If they really wanted to fight disinformation, they would do this, but clearly they don't want to. They are necessary co-operators in the dissemination of false news. (Interview)

Macías-Tovar also adds that it is essential that all fact-checkers be transparent and show two crucial aspects in a clear and open way:

> Although they do a good job, I have two drawbacks to highlight: one is that they do not say who finances these campaigns, and the second is that they do not say that those who lie today are those who lied yesterday, the day before yesterday and also four days ago. I don't think they go into much depth there and that is where I am trying to move forward. (Interview)

Despite his work, Macías-Tovar has a pessimistic view of fact-checking as he considers the fight against disinformation is presented as an almost impossible and difficult task to accomplish.

> You lie in one sentence in just ten seconds and it has a tremendous viralisation. To do a good fact-checking you have to research for days and make a publication that can be read for several minutes; therefore, in communicative engineering the lie always wins. (Interview)

Analysing his experience with the fact-checker he found that the virality of the dissemination of fake news and the extension of digital noise is based

on the most well-known advertising techniques applied to networks. They tend to permeate very easily, says Macías-Tovar, because it is something that seems to be very well studied and has a very powerful psychological component that manages to "hook" people. This is inspired by the classic principles of advertising but applied to the current society of spectacle where the image, the emotional story and the framing set the trend (Díez-Gracia & Sánchez-Gracia, 2022; Durán, 2021; Marcos, 2018; Muñiz & Navazo, 2021).

According to Macías-Tovar, in order to curb the circulation of fake news through networks and in the media, it would be necessary to come up with different measures:

> Some kind of legislative consensus would have to be reached. Fake news should not be allowed to run free without anybody to stop it. It would be important for this consensus not to come from a specific government, but from a network of journalists, jurists, experts, etc. (Interview)

To do this it is essential to get rid of lies and a solution would be to penalise the media that spread lies:

> What is true and what is a lie may sometimes be debatable; however, outright lies need to be kept out of the media and the media outlets that spread lies need to be penalised. (Interview)

Thus, it is necessary to recognize that although fact-checking is not always capable of repairing the damage caused by fake news, as Macías-Tovar argued, its very existence can help to reduce the disinformation spread (Nyhan & Reifler, 2012). Moreover, 38.5% of fact-checkers (Dafonte-Gómez et al., 2019) also play the role of "educators" by accompanying their work of verification, transparency, and critical analysis with training strategies for media audience literacy, the identification of false news and the behaviour in response to them.

Some questions for reflection

The study of the campaign on macro-farms shows that the use of fake news on social networks goes beyond disinformation. Its intention is also to influence the perception of reality and the construction of a polarised socio-political climate that can have very serious effects on offline reality (Müller & Schwarz, 2021), as we have seen. The political use of fake news also plays a very relevant role since it is used by opposition parties; i.e. the objective is not only to promote digital noise, but to provoke conflict in the real world.

We are not talking about a possible offensive or unpopular discourse which logically is permitted and protected by freedom of expression (Martínez--Torrón, 2016), but about a form of manipulation of reality and information that can undermine not only the right to truthful information but also democratic deliberation on the network and democracy itself (Díez-Gutiérrez et al., 2022).

Indeed, research reveals how without the participation and the vehicle of traditional media (television, newspapers, radio, etc.) campaigns would not have the high level of impact they achieve or the effects they entail (Hernández-Conde & Fernández-García, 2019).

In social networks it seems that the political hoax that seeks to provoke emotional reactions such as ideological hatred, confrontation and polarisation is easily transformable into 'online social capital', transformed into an increase in followers, more retweets, quotes or likes, in an environment in which the goal has become virality and this seems to be the culmination of 'digital success'.

One of the strategies of disinformation campaigns is to achieve a large reaction to fake news or messages of noise and hatred (Magallón-Rosa & Campos, 2021). In the same sense, post-truth is related to those processes where reality is explained emotionally and not by empirically demonstrated facts (Aparici & García-Marín, 2019). In the digital era these "battles" on

Fact-checking on Twitter in the face of fake news:
Combating hate speech on social networks

social networks and the media in coordination with political actions and those of foundations or social groups have been called digital hybrid wars (Polanco, 2021).

Twitter is a company that seeks to increase profits for its shareholders. Avoiding the growth of polarisation by this platform would go against this business model that reinforces its interaction and expansion (Domínguez, 2021). Therefore, although Twitter repeatedly announces that it is going to strengthen its verification and control filters and that it is developing mechanisms to improve them in order to reduce false information, it still has a high proportion of fake news on its network (Popiolek et al., 2021).

The big technology multinationals have a primary interest: to promote those messages, tweets, retweets, controversies that involve more likes and more interactions. This implies higher profits for them which is the key element of their business model (Estévez, 2022). Far from being random, the algorithm benefits hatred much more frequently compared to other types of emotions, according to Macías-Tovar and confirmed by research (Magallón--Rosa & Campos, 2021). In this sense, it is essential and also an ethical issue to establish transparency standards regarding the algorithms and automatic procedures of social networks. There is a need for greater public control over the technological giants also for the sake of greater transparency and control over the algorithms behind social networks. In the context of this research and following the interview with the head of Pandemia Digital (Digital Pandemic) we propose possible ways of intervening in the face of disinformation:

It is necessary to combat disinformation and its effects on citizens. One of the most important measures would be to prevent those who spread lies from having total impunity. Political will is essential to achieve this. In this regard we lack a legislative model adapted to new scenarios, formats, actors, and new challenges seeking the defence of pluralism of information (Magallón-Rosa, 2019). In this sense: "any initiative needs to distinguish between legislative, digital literacy, training and action measures, and

between internal and external measures, knowing their updating may be necessary with each election" (Magallón-Rosa, 2019: 345).

Likewise, citizen collaboration is necessary to undertake critical work, to question and prevent this type of digital noise (Mateus, Andrada, & Ferrés, 2019), and to help professional verifiers such as fact-checkers to stop and prevent it. Thus, it is necessary to recognise that although verification does not always repair the damage caused by fake news, it certainly contributes to reducing disinformation.

Finally, from an educational approach we also consider it necessary to act preventively against disinformation by promoting media information literacy and critical *educommunication* (Fernández-García, 2017). Poor critical digital competence is one of the key factors most highlighted in the academic literature (Baptista & Gradim, 2020) influencing the consumption and distribution of disinformation. Educational spaces need to promote media literacy focused on knowledge and the development of skills that favour the critical understanding and conscious use of digital media (González, Tucho, & Marfil-Carmona, 2020; Sierra & Dafonte-Gómez, 2021), in both formal and informal education (Nieto, 2021).

As limits of this research, we understand that a single interview is perceived as scarce; it would be appropriate to conduct more interviews with professionals and their subsequent analysis to find elements in common. In this same sense we must also point out the restriction in the generalization of the results, considering the sample as reduced and circumscribed to a social network.

References

Almagro, E. & Cabrera, S. (2022, February 1). The violent assault on the plenary in Lorca told by his witnesses: "We never thought they would reach this extreme". *The Diario.es*. https://cutt.ly/kOQUbPW

Alonso, M. (2019). Fake news: Disinformation in the age of the information society. Areas. *International Journal of Communication*, *45*, 29-52. https://doi.org/10.12795/Ambitos.2019.i45.03

Amorós, M. (2018). *Fake news: The truth of fake news.* Current Platform.

Aparici, R. & García-Marín, D. (2019). *Post-truth. A mapping of media, networks and politics.* Gedisa.

Baptista, J. P. & Gradim, A. (2020). Understanding fake news consumption: A review. *Social Sciences, 9*(10), 185. https://doi.org/10.3390/socsci9100185

Briceño-Montilla, L. A. (2022). Capitalism of seduction, exposure and emotion in the gradual society of 'likes'. *Humanities, 12*(1). http://dx.doi.org/10.15517/h.v12i1.49300

Calvo, D., Campos-Domínguez, E., & Díez-Garrido, M. (2019). Towards a computational campaign: Online tools and strategies in the Spanish elections. *Revista Española de Ciencia Política,* 51, 123-154. https://doi.org/10.21308/recp.51.05

Calvo, E. & Reche, E. (2022). This is how the assault on the plenary session of Lorca was conceived. *The Diario.es. https://cutt.ly/KOTwbKQ*

Carvajal, M., Mondéjar, D., Valero-Pastor, J. M., De-Lara, A., García-Avilés, J. A., & Arias-Robles, F. (2022). The most outstanding journalistic innovations in Spain (2010-2020): Characteristics and organizational, industrial and social impact. *Information Professional*, *3*(3). https://doi.org/10.3145/epi.2022.may.04

Casero-Ripollés, A. (2020). Influence of media on the political conversation on Twitter: Activity, popularity, and authority in the digital debate in Spain, *Icono14, 18*(1), 33-57. https://doi.org/10.7195/ri14.v18i1.1527

Castro, R. (2021). Participant observation. Stories of a life and in-depth interview. In: J. M. Tejero (Ed.), *Qualitative research techniques in the health and social health fields* (pp. 85-102). Editions of the University of Castilla-La Mancha.

Córdoba-Cabús, A. & López-Martín, Á. (2021). The use of Twitter as the main dissemination channel for data journalism. In: J. Sotelo González & J. González García (Coords.), *Digital media* (pp. 225-244). McGraw Hill.

Dafonte-Gómez, A., Corbacho-Valencia, J. M., & García-Mirón, S. (2021). Fact-checking in Latin America: Recent developments and situation map. In: J. Sotelo González & J. González García (Eds.), *Digital Media* (pp. 877-889). McGraw-Hill.

Díez-Gracía, A. & Sánchez-García, P. (2022). The news gap in the "triple digital agenda": The different interests of media, audience and networks. *Communication & Society, 35*(1), 63-80. https://doi.org/10.15581/003.35.1.63-80

Díez-Gutiérrez, E., Verdeja, M., Sarrión-Andaluz, J., Buendía, L., & Macías-Tovar, J. (2022). Political hate speech of the far right on Twitter in Latin America. [Far-right political hate speech from Twitter in Latin America]. *Communicate, 72,* 101-113. https://doi.org/10.3916/C72-2022-08

Digital 2021. (2021). *We Are Social*. https://bit.ly/3bQqtvD

Dominguez, D. (2021, April 15). Twitter, a hate network? *The Tide*. https://cutt.ly/1OTZGKw

Durán, D. D. (2021). Reality in 280 characters. The legitimized lie. *Revista Mexicana de Comunicación,* 146-147.

Ecker, U. H., O'Reilly, Z., Reid, J., & Chang, E. (2019). The effectiveness of short-format refutational fact-checks. *British Journal of Psychology,* 111, 36-54. https://doi.org/10.1111/bjop.1238

Elias, C. (2021). Journalism as a tool against fake news. In: C. Elías & D. Teira (Eds.), *Manual de periodismo y verificación de noticias en la era de las fake news* (pp. 19-57). Editorial UNED.

Estévez, J. A. (2022). The algorithmization in the world of surveillance capitalism. *OXÍMORA International Journal of Ethics and Politics*, (20), 1–37. https://doi.org/10.1344/oxi.2022.i20.38108

European Commission (2018). *A multi-dimensional approach to disinformation. Report of the high level group on fake news and online disinformation*. https://bit.ly/3oLtcM1

Fernández-García, N. (2017). Fake news: An opportunity for media literacy. *New Society,* 269, 66-77.

García-Crespo, O., Ramahí-García, D., & Dafonte-Gómez, A. (2021). Ibero-American fact-checkers on Youtube. Analysis of presence and activity. In: M. Blanco Pérez (Coord.), *The progress of communication in the age of prosumers* (pp. 83-106). Dykinson.

García-Marín, D. (2022). Algorithmic models and automated fact-checking. Systematic review of the literature. *Documentation of Information Sciences, 45*(1), 7-16. https://dx.doi.org/10.5209/dcin.77472

Gómez-Calderón, B. (2020). The phenomenon of "fake news" or how new channels serve as a loudspeaker to the old propaganda. In: F. Martín & M. Gómez (Eds.), *Post-periodismo. Between the real and the virtual* (pp. 21-42). McGraw Hill Inter-American.

González, D. A., Tucho, F., & Marfil-Carmona, R. (2020). The dimensions of media competence in Spanish university students. *Icon14, 18*(2), 217-244. https://doi.org/10.7195/RI14.V18I2.1492

Graves, L. (2016). Anatomy of a fact check: Objective practice and the contested epistemology of fact checking. *Communication, Culture & Critique, 10*(3), 518–537. https://doi.org/10.1111/cccr.12163

Graves, L. (2018). Boundaries not drawn: Mapping the institutional roots of the global fact checking movement. *Journalism Studies, 19*(5), 613–631. https://doi.org/10.1080/1461670X.2016.1196602

Guallar, J., Codina, L., Freixa, P., & Pérez-Montoro, M. (2020). Misinformation, hoaxes, healing and verification. Review of studies in Ibero-America 2017-2020. *Telos, 22*(3), 595-613. https://doi.org/10.36390/telos223.09

Hernández-Conde, M. & Fernández-García, M. (2019). Emerging parties of the far right: Fake news, fake outsiders? *Teknokultura, 16*(1), 33-53. https://doi.org/10.5209/TEKN.63113

Humprecht, E. (2019). How do they debunk "fake news"? A cross-national comparison of transparency in fact checks. *Digital Journalism, 8*(3), 310-327. https://doi.org/10.1080/21670811.2019.1691031

Lim, Ch. (2018). Checking how fact-checkers check. *Research & Politics,* 1-17. https://doi.org/10.1177/2053168018786848

López-Pan, F. & Rodríguez, J. (2020). Fact checking in Spain. Platforms, practices and distinctive features. *Studies on the Journalistic Message, 26*(3), 1045-1065.

Losada, R. (2020). Fake news in the crises of the digital society. *More Local Power,* 41, 48-50.

Luque, T., Macías-Tovar, J., & Casado, F. (2021). *The democracy of the trolls and the assault on democracy. Analysis of the electoral process in Ecuador and the radiating elements that make it up through its social networks.* UTM Editions.

Macías-Tovar, J. (2022). Alberto Garzón's victory in three acts: The lie, the truth and the ridicule. *Public.* https://cutt.ly/UPbWntP

Magallón-Rosa, R. (2019). The (non)regulation of disinformation in the European Union. A comparative perspective. *Journal of Political Law, 1*(106), 319-346. https://doi.org/10.5944/rdp.106.2019.26159

Magallón-Rosa, R. & Campos, E. (2021). Social networks and polarization. When the algorithm amplifies human emotions. *Dossieres esf,* (42), 27-31.

Marcos, S. (2018). *Social networks as a tool for political communication. Political and citizen uses of Twitter and Instagram* (Doctoral dissertation, Universitat Jaume I).

Martínez-Torrón, J. (2016). Freedom of expression and offensive language: Some practical criteria for legal analysis. *The Chronicler of the Social and Democratic State of Law,* (60), 26-33.

Mateus, J. C., Andrada, P., & Ferrés, J. (2019). Assessing media competence: A critical approach from the pedagogical, political and methodological perspectives. *Journal of Communication, 18*(2), 287-301. https://doi.org/10.26441/RC18.2-2019-A14

Müller, K. & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association, 19*(4), 2131–2167. https://doi.org/10.1093/jeea/jvaa045

Muñiz, J. A., & Navazo, P. (2021). Visual disinformation in social networks and digital media: A taxonomic proposal of distributed photographic manipulation during the pandemic. *Information and Communication Notebooks, 26,* 77-87. https://doi.org/10.5209/ciyc.75235

Nieto, M. I. (2021). The EU against disinformation. In: M. Blanco-Pérez (Coord.), *The progress of communication in the age of prosumers* (pp. 137-148). Dykinson.

Nyhan, B. & Reifler, J. (2012, February 28). Misinformation and fact-checking: Research findings from social science. *New America.* https://bit.ly/3wyFmdU

Jamison, A. M., Broniatowski, D. A., & Quinn, S. (2019). Malicious actors on Twitter: A guide for public health researchers. *American Journal of Public Health, 109*(5), 688-692. https://doi.org/10.2105/AJPH.2019.304969

Janssen-Observer-FAPE-National Association of Health Informants (ANIS) (2017). *II Janssen Observer Poll Journalists and Social Networks.* https://bit.ly/3nPkegt

Otero, J. (2022, February 1). Lorca: X-ray of the assault on popular sovereignty by the macrofarms encouraged by the right and the ultra-right. *Public.* https://cutt.ly/ROQnDyd

Paz-García, A. P. & Spinosa, M. M. (2014). Political journalism on Twitter: Electoral coverage in #CBAVota (Argentina, 2011). *Cuadernos.info,* (34), 133-152. https://doi.org/10.7764/cdi.34.558

Polanco, R. (2021). The new instruments of the hybrid war of the United States against Venezuela in the digital public space, in a context of continuous coup. *International Political Review, 3*(4), 131-147.

Popiołek, M., Hapek, M., & Barańska, M. (2021). Infodemic. *Communication and Society, 34*(4), 81-98.

Reuters Institute. (2020). *Digital News Report 2020.* https://bit.ly/36Uw7um

Reuters Institute. (2019). *Digital News Report 2019.* https://bit.ly/39h4iOi

Salaverría, R., Buslón, N., López-Pan, F., León, B., López-Goñi, I., & Erviti, M. C. (2020). Disinformation in times of pandemic: Typology of hoaxes about Covid-19. *The Information Professional*, 29(3). https://doi.org/10.3145/epi.2020.may.15

Sharevski, F., Jachim, P., & Florek, K. (2020). To tweet or not to tweet: Covertly manipulating a Twitter debate on vaccines using malware-induced misperceptions. *ArXiv*, 1. Cornell University. https://bit.ly/3a4dvbV

Sierra, S. & Dafonte-Gómez, A. (2021). Media literacy resources and training initiatives: A review of Ibero-American fact-checkers. In: M. Blanco-Pérez (Coord.), *The progress of communication in the age of prosumers* (pp. 217-238). Dykinson.

Singer, J. (2018). Fact-checkers as entrepreneurs: Scalability and sustainability for a new form of watchdog journalism. *Journalism Practice, 12*(8), 1070-1080. https://doi.org/10.1080/17512786.2018.1493946

Ufarte-Ruiz, M., Peralta-García, L., & Murcia-Verdú, F. (2018). Fact checking: A new challenge of journalism. *The Information Professional*, 27(4), 733-741. https://doi.org/10.3145/epi.2018.jul.02

Vizoso, A. & Vázquez-Herrero, J. (2019). Fact-checking platforms in Spanish. Characteristics, organization and method. *Communication and Society, 32*(1), 127-144. https://doi.org/10.15581/003.32.1.127-144

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Young, D., Jamieson, K., Poulsen, S., & Goldring, A. (2018). Fact-checking effectiveness as a function of format and tone: evaluating FactCheck.org and FlackCheck.org. *Journalism and Mass Communication Quarterly, 95*(1), 49-75. https://doi.org/10.1177/1077699017710453

Authors

EDITOR

Branco Di Fátima is a non-fiction writer with a PhD in Communication Sciences from the University Institute of Lisbon (ISCTE). He wrote the book *Dias de Tormenta* (Geração Editorial, 2019) and edited the collection *Hate Speech on Social Media: A Global Approach* (LabCom Books, EdiPUCE, 2023). He has published more than 90 scientific works and participated in 11 research projects funded by national and international organizations. His current research focuses on the pathologies and dysfunctions of democracy, journalism studies, online hate speech, and social network analysis. He is currently a contracted researcher at LabCom – University of Beira Interior (UBI) in Portugal.

AUTHORS

Adolfo Carratalá is Associate Professor of Journalism at the University of Valencia. Previously, he was a post-doctoral researcher at the Rey Juan Carlos University (URJC) in Madrid and Associate Professor of Communication at the International University of La Rioja (UNIR). He holds a PhD in Journalism. His PhD thesis was awarded by the Spanish Society of Journalism. His primary research interests are communication and conflicts, media and social movements, framing and LGBTIQ+ Media Studies.

Daniel Martin Pena is PhD in Communication and Education, Associate Professor at the Faculty of Documentation and Communication Sciences of the University of Extremadura Spain and researcher in radio and university media and scientific dissemination.

Debora Cristina Lopez holds a PhD in Contemporary Communication and Culture from the Federal University of Bahia (Brazil) and carried out a post-doctoral internship at the University of Extremadura (Spain). She is a professor in the Postgraduate Program in Communication and the Undergraduate Program in Journalism at the Federal University of Ouro Preto (Brazil). She has a CNPq productivity grant. She coordinates the Convergence and Journalism Research Group (ConJor).

Enrique Javier Díez-Gutiérrez. Professor of the Faculty of Education at the University of León (Spain). PhD in Educational Sciences. He develops his teaching and research work in the field of educational policy, intercultural education, and equality. Director of the European Research Erasmus +, Jean Monnet Module "Building up an Inclusive and Democratic Europe through a Dialogical Co-Creation of Intercultural Solutions to the Rise of Neo-Fascism and Xenophobia". EAC-A02-2019-JMO. Application 620320-EPP-1-2020-1-ES-EPPJMO-MODULE. Among his latest publications are Antifascist Pedagogy. Construir una educación inclusiva, democrática y del bien común frente al auge del fascismo y la xenofobia (Octaedro, 2022). The silenced history. Unidades Didácticas de Recuperación de la Memoria Histórica Democrática (Plaza y Valdés, 2022). Critical and inclusive education for a post-capitalist society (Octaedro, 2021). La asignatura pendiente. La memoria histórica en los libros de texto (Plaza y Valdés, 2020), La educación en venta (Octaedro, 2020), Educación para el bien común (Octaedro, 2020), La revuelta educativa neocon (Trea, 2019) or Neoliberalismo educativo (Octaedro, 2018).

Eugenia Siapera is Professor of Information and Communication Studies and the co-Director of the UCD Centre for Digital Policy (with Elizabeth Farries). Her research interests are in the area of digital platforms, political communication and journalism, technology and social justice, platform governance and hate speech, racism and misogyny. She was the PI of an IRC-funded project on racist hate speech in the Irish digital sphere, and a partner in a H2020 project on the mediated memory of conflict (RePAST). She is the PI of a research project on Alt Tech Platforms and Harmful

Narratives. She has written numerous articles and book chapters. Some of her recent book projects include Understanding New Media (Sage, 2018, second edition), Gender Hate Online (2019, Palgrave, co-edited with Debbie Ging) and Radical Journalism (2023, Routledge co-edited with Seamus Farrell and George Souvlis). She is currently working on the third edition of Understanding New Media and on an edited volume on Cancel Culture (with Paraic Kerrigan and Elizabeth Farries, under contract with Routledge).

Eva Palomo-Cermeño. She holds a PhD from the Universidad Rey Juan Carlos (URJC) where she works as a lecturer in the area of Moral Philosophy. She holds a Master's degree in Interdisciplinary Gender Studies. Her lines of research are: Applied ethics, women's studies, feminist theory, and gender-based violence. Besides articles published in these fields, in 2015 she has written the book Sylvia Pankhurst. Suffragette and socialist (Ed. Almud). The last project she has participated in as a researcher is "An interdisciplinary analysis of the sources of legitimacy and normalization of the new pornography industry" (URJC).

Imran ul Haq is currently a doctoral scholar at the Department of Media Studies, Bahria University, Islamabad. He pursued studies in English Literature and Mass Communication before embarking on his career as a lecturer. He brings experience from renowned organizations such as the European Press Agency and presently holds the position of Senior Publications Officer at the Pakistan Institute of Development Economics. His academic pursuits and professional journey reflect a keen interest in Political communication, media accountability and self-regulation.

Jamal ud Din is associated with University of Swat. He has been a journalist and worked with the prestigious media organizations in Pakistan. He regularly publishes on topics related to political communication and journalism practices in Pakistan.

Lina Uribe-Henao is a communicator from the Pontificia Universidad Javeriana Cali and a Master in Colombian and Latin American Literature from the Universidad del Valle. She has worked as a journalist, teacher,

and consultant. She currently works as Communications Director of the Observatory for Women's Equity (OEM), as a lecturer in the Department of Communication in Digital Environments at Universidad Icesi (Cali, Colombia), and as a lecturer in the Department of Communication and Language at Universidad Javeriana Cali.

Marcelo Freire holds a PhD in Contemporary Communication and Culture from the Federal University of Bahia (Brazil) and carried out a post-doctoral internship at the University of Extremadura (Spain). He is a professor in the Postgraduate Program in Communication and the Undergraduate Program in Journalism at the Federal University of Ouro Preto (Brazil). He coordinates the Convergence and Journalism Research Group (ConJor).

Maria Juliana Soto-Narváez is a teacher and researcher. She has worked in the field of human rights concerning technology, communication, and digital culture. Currently, she is the head of the Department of Communication in Digital Environments at Universidad Icesi (Cali, Colombia). She holds a Master's degree in Communication and Media from the Universidad Nacional de Colombia and is a Social Communicator-Journalist from the Universidad del Valle.

María Verdeja Muñiz. University of Oviedo. Associate Professor. Faculty of Teacher Training and Education. Area of Didactics and School Organisation. PhD in Pedagogy. Her main lines of research focus on the study of the pedagogy and work of Paulo Freire, educational policy, the challenges posed by education in multicultural contexts, intercultural education, and attention to the cultural diversity of students from a focus on inclusive education and social justice. As a result of this research, he has published in scientific journals and has been a speaker at national and international conferences.

Mauricio Guerrero-Caicedo is a sociologist from Universidad del Valle with a master's degree in Journalism from Universidad Icesi. His academic interests are related to the production of content in the digital environment, the processes of creation in virtual communities, and the implications of

disinformation. During his professional career, he has worked as a teacher, researcher, consultant, and creative. He currently serves as director of the Communication program at Universidad Icesi.

Sarah Anne Dunne is based in UCD School of Information and Communication Studies. She is currently postdoctoral researcher and administrator for UCD Centre for Digital Policy. She has previously acted as a postdoc on the IRC funded project Platforming Harm, which analyses the circulation of harmful health narratives and far-right ideology on Alt-Tech platforms. Her doctoral research focused on the dissemination of rape culture, feminist activism and misogyny on social networking platforms. Her research interests include feminist theory, gender and sexuality studies, critical race theory and digital cultures.

Sergio G. Grabosky is a Language Professor in secondary and higher education at the National University of Salta, Argentina. He specializes in Social Sciences with an emphasis on Reading, Writing, and Education. He holds a degree from FLACSO, Costa Rica, and a master's degree in Social Sciences (Master's Degree II CIREF) from the Université de Reims, Champagne-Ardenne, France. He is an Associate Professor in the Licentiate program in Communication Sciences at the School of Humanities, National University of Salta, Argentina, and a Professor in the master's program in Contemporary Issues of Communication at the Postgraduate Department of the School of Humanities and Social Sciences, National University of Jujuy, Argentina. He leads Research Project No. 2724/0 Writing in Communication Science for the Research Committee of the National University of Salta (CIUNSA) and is part of RED Investigar en Red, coordinated by Ph.D. Albarello Francisco.

Shabir Hussain holds the position of Professor in the Department of Media Studies at Bahria University, Islamabad. His research mainly focusses on problematic information in politics and foreign policy arenas.
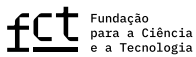
This is the first book of the **Online Hate Speech Trilogy**. It focuses on the relationship between disinformation, political polarisation, and virtual attacks involving toxic language. These phenomena are much more complex than they may initially appear, and they are strongly interrelated with each other in the algorithmic society. The authors analyse the deep origins of hate speech and its manifestations online. They highlight attacks against the LGBTQ+ community in Europe, journalists in Asia, and black politicians in Latin America. Additionally, these researchers examine the crucial role of fact-checking agencies in confronting toxic language and combating hate speech.