

Social networks, monitoring equipment, and mobile apps to combat hate speech and hate crimes in Europe

Redes, equipos de monitoreo y aplicaciones móvil para combatir los discursos y delitos de odio en Europa

Roberto Moreno López.

Universidad de Castilla-La Mancha. Spain.

roberto.moreno@uclm.es



César Arroyo López.

Instituto de estudios internacionales en educación y sociedad. Spain.

cesar.arroyo@ineeys.es



How to cite this article / Standard reference

Moreno López, R. & Arroyo López, C. (2022). Social networks, monitoring equipment, and mobile apps to combat hate speech and hate crimes in Europe. *Revista Latina de Comunicación Social*, 80, 347-363. <https://www.doi.org/10.4185/RLCS-2022-1750>

RESUMEN

En este contexto global, inmediato y participativo, con carácter comunicativo propio, se ha visto en los últimos años como un espacio de expresión y propagación de la intolerancia. Con este caldo de cultivo y utilizando el mencionado anonimato como medio para la impunidad, surge el ciberodio. En el marco del proyecto C.O.N.T.A.C.T. se desarrolla una investigación de tipo mixto con cuestionario ad-hoc (N=186) y 20 entrevistas en profundidad a jóvenes entre los 18 y los 35 años en España con el objetivo central de profundizar en la percepción y los conocimientos sobre discurso de odio presente en la juventud española. Los principales resultados arrojan que el discurso de odio forma parte de la comunicación en internet y las redes sociales y que el alcance e impacto de los discursos de odio se retroalimentan tanto en la calle como en la red. Los jóvenes entienden que el discurso de odio se utiliza contra determinados grupos sociales por su orientación física, cultural, étnica, sexual y aunque pueden llegar a reconocer su uso en el ámbito privado. También identifican el discurso de odio como una estrategia de intimidación. En definitiva, la lucha contra el odio en internet es una cuestión multifacética, donde las medidas de carácter legal se deben acompañar de propuestas y respuestas de carácter social y educativo, y debe contar con el compromiso institucional y de las empresas que prestan servicios tecnológicos, todo ello orientado a construir un espacio online más seguro y respetuoso para todos.

PALABRAS CLAVE: Redes sociales; discursos de odio; delitos de odio; jóvenes; TICs

ABSTRACT

In this global, immediate, and participatory context, with its own communicative character, it has been seen in recent years as a space for the expression and propagation of intolerance. With this breeding ground and using the anonymity as a means for impunity, arises cyber-hate. Within the framework of the C.O.N.T.A.C.T. project, qualitative-quantitative research was carried out with an ad-hoc questionnaire (N=186) and 20 in-depth interviews with young people between 18 and 35 years of age in Spain with the main objective of deepening the perception and knowledge of hate speech among Spanish youth. The main results show that hate speech is part of communication on the Internet and social networks and that the reach and impact of hate speech feeds back both on the street and online. Young people understand that hate speech is used against social groups determined by their physical, cultural, ethnic, sexual orientation and although they may come to recognize that they use it in the private sphere. They also identify hate speech as an intimidation strategy. In short, the fight against hate on the Internet is a multifaceted issue, where legal measures must be accompanied by proposals and responses of a social and educational nature and must have the commitment of institutions and companies that provide technological services, all aimed at building a safer and more respectful online space for all.

KEYWORDS: Social media; hate speech; hate crimes; young people; TICs

CONTENIDO

1. Introducción. 2. Objetivos. 3. Metodología. 4. Discusión/Resultados 5. Conclusiones. 6. Bibliografía 7. Curriculum Vitae

CONTENIDO

1. Introduction. 2. Objectives. 3. Methodology. 4. Discussion/Results 5. Conclusions. 6. Bibliography 7. Curriculum Vitae

Translation by **Paula González** (Universidad Católica Andrés Bello, Venezuela)

1. Introduction

The technological revolution that began at the end of the 1960s and the expansion of the Internet since the end of the 1990s represented a real shock that has not yet stopped. Its depth was so deep that it completely moved and affected all areas of society. Its impact on the economy, culture, and communication are undeniable. We have gone from a unilateral communication dominated by the big media and administrations to a knowledge society where citizens can not only become transmitters of information but also co-creators of new content in a previously much more limited way. In this internet world, as a space for global relationships where space-time limitations have been diluted (Sacristán, 2013, p.133), anyone with access to the resources that allow them to connect to the online world can offer their opinion, contribute to dialogue and global discussion, and build knowledge with their perception and ideas, contributing to modern cultural gestation or "cyberculture" (Sacristán, 2013, p.126).

Authors such as López & Sánchez (2019) raise the importance of digital coexistence by presenting attitudes towards the virtual and digital worlds. Attitude is a key element that is used to interact with other users. Through the development and identification of this attitude, we can locate possible risks of violating intimacy and privacy in the way we interact with each device. For authors such as De Haro (2019), social networks at this time can form other more general independent networks that isolate them from other Internet users and create safe spaces to catalyze different human behaviors and provide new communication and learning formats (Holcomb & Bill, 2010). On the other hand, the rise of social networks in our society implies not only a change in traditional communicative referents seen

in general terms but also a growing need to be aware of how scientific actors and agents use these challenges.

It should be noted that, in this online framework, communication on the internet and social networks is endowed with a series of particularities, among which we will especially highlight the experience of anonymity and the feeling of disinhibiting virtuality (Barcelona City Council, 2017). So, in this global, immediate, and participatory context, with its own communicative characteristics, it has been perceived over the last few years that it is becoming a space for the expression and propagation of intolerance (Cabo & Juanatey, 2016). An intolerance that underlies the non-acceptance of what is different, which is intended to strip the person of their dignity by denying and attacking their identity. Furthermore, it is known that this intolerance on the Internet "not only negatively affects the groups or individuals to whom it is directed, but also those who defend freedom, tolerance, and non-discrimination" (European Commission, 2016, p. 1). As social networks have come to dominate the sociopolitical landscape in almost every corner of the world, more and more racist acts old and new are taking place on these platforms (Matamoros-Fernández & Farkas, 2021). Racist rhetoric thrives on social media, including covert tactics such as militarization through memes (Lamerichs et al., 2018) and the use of false identities to incite racist hate (Farkas et al., 2018).

Although inappropriate speech is prohibited by international law and regulatory policies based on respect for humanity, it has become widespread, endangering the values necessary for social cohesion and, in some cases, hate speech can fuel tensions and incite violence. It can be addressed to one person, to a group of people, or to no one in particular (Davidson et al., 2017), which makes it a difficult to define and multidimensional problem. In Europe in particular, as part of the global North, hate speech is seeping into public discourse, especially in the wake of the refugee crisis, which was largely unleashed in 2015 (Ekman, 2019). In this sense, its impact in real life is also increasing, since it can be a precursor and trigger of hate crimes (Burnap & Williams, 2014). Many people are quickly recognizing that hate speech is a serious problem, especially through social media sites, leading to more cyber conflicts between different people (Al Serhan et al., 2020).

Social stereotypes fuel hate speech, found in real life and online, and discussions have recently begun to revolve around providing unruly free speech and, in some cases, uncontrolled hate speech through digital technologies, and social networks have even developed their own services to detect and prohibit this type of rhetorical expression (Pohjonen & Udupa, 2017) which, despite the official policies of the platforms in their conditions of service, are covert or overt (Ben-David & Fernández, 2016). Of course, the distinction between hate and offensive speech must be made clear, and this process is aided by the definition of legal terminology. Mechanisms that control and analyze abusive language are established in efforts to recognize the aggressive speech that expands in the online media, to the extent that its technological possibilities allow. Following Olteanu, et. al (2018) the spread of hate sentiment has intrigued many researchers who initially research online content to aid surveillance and, after analyzing the results, make it easier for politicians, and policymakers to understand and find solutions in a contextualized framework.

With this breeding ground and using the aforementioned anonymity as a means of impunity, cyber-hate is understood (in a non-restrictive way) as any use of electronic communications technology to spread anti-Semitic, racist, intolerant, extremist, or terrorist messages or information. These electronic communications technologies include the Internet (Anti-Defamation League, 2010) as an extension of hate speech to the digital sphere and anticipation of hate crimes, finding the ideal space for its propagation and amplification on the Internet. On the other hand, Paz et al. (2020) refer to four factors, means used to spread hate speech, the subject of the speech, the area in which the speech is produced, and the roots or novelty of the phenomenon and its evolution, which each offers variables of quantification

and qualification that must be exploited through various methodologies and interdisciplinarity. Therefore, hate violence is considered to be the result of hate incidents and hate crimes that negatively impact the physical and mental health of victims, both in the short and long term (Gil-Borrelli et al., 2018; Gil-Borrelli et al., 2020). When expressed in this way, it is a term that is easier to understand and relevant to the healthcare context. In this context, the impact of hate violence is very important, since it can have physical and psychological effects, as well as change people's behavior, and even have an impact on group integration into the community to which the victim belongs.

In Spain, as detailed in the latest report on the evolution of incidents related to hate crimes in Spain (Ministry of the Interior, 2020), 1,334 acts related to hate crimes were recorded for that year, of which 189 were made through the internet or social networks. 45% through the internet and 23% through social networks. For its part, the Proxi Observatory, through the Online Project against Xenophobia and Intolerance, analyzed almost 5,000 comments in three large digital newspapers and concluded that more than half of the comments that appeared associated with the news in these digital newspapers were of an intolerant nature (Cabo, Tarragona, & Vallès, 2015). There are also in our country, cases of online hate of great relevance and social impact of the use of the Internet and social networks for incitement to violence and hatred against people based on racism (Moha Gerehou case) or homophobia (Jesús Tomillero case), among others. These data, by way of example, confirm the generalized view that intolerance and hatred have special accommodation and diffusion on the Internet.

What role does anonymity play in the spread of bigotry and hate speech on social media, then? The Internet is a space that provides its users with the capacity for expression and communication without limits, without State control, or without the pre-eminence of pressure groups and the media, as it makes identification difficult (when not totally hidden), as well as the location and activity of the user. According to De Salvador (2012), the concept of anonymity on the Internet is the ability to carry out any access, communication, or publication on the Internet without third parties having the possibility of identifying or locating the author of the action. Although this anonymization is indeed feasible through various strategies and tools, most of the public communication that occurs is traceable at the origin. In the case at hand, that of young users who communicate via the Internet, there was a perception among those interviewed that much of the communication that occurs on the Internet is anonymous, for the simple fact that it is not done face-to-face, and that this concealment allowed any type of behavior.

ICTs and internet access are a defining and present element in the lives of young people. Only in Spain, the use of the Internet daily among young people between 16 and 35 years of age is above 98% (INE, 2021). Therefore, the presence and interaction of young people in the online context and the use of devices for online life are constant. The experience, however, is not always positive, and as some recent studies point out, traditional bullying has leaped to the digital world, to the point that the number of cases practically equals both (Save the Children Spain, 2016) and where the insult is the main manifestation of internet harassment (Romera et. al., 2017).

Before looking for the possible links between the phenomenon of bullying, cyberbullying, and hate speech, it is necessary to make a conceptual approach to the terms. Bullying is a social phenomenon of great relevance, which consists of "a type of behavior aimed at harming, repeated over time, and that occurs within an interpersonal relationship characterized by an imbalance of power" (Olweus, 1999). For its part, cyberbullying has been considered by some authors as part of a series of violent behaviors that occur through ICTs. Hate speech, for its part, is similar to bullying in that both are aggressive behaviors, of an intentional nature, that are repeated over time, with an imbalance of power, but with the particularity that the person or persons who undertake these behaviors do so through electronic means by which they contact the victim. As mentioned by Del Río et al. (2010) as a form of harassment, cyberbullying has some particularities that make it especially intense, such as its great scope

concerning the audience that observes it, the perception of anonymity of the harasser, the absence of time limits, the imperishability of the published content, the speed with which the message is spread, and the change in the elements that marked the power relationships in bullying. In short, cyberhate, which is linked to the concept of hate speech (which is enough to take a look at the different international documents, does not have a universally agreed definition), can be considered as the use of electronic communication technologies for the dissemination of intolerant, discriminatory information or messages or that encourage and promote violence against certain people or groups based on their identity. It also has its specificities, some of them shared with cyberbullying, such as: communicative overabundance, the possibility that anyone can send hateful messages, the possibilities of multiplying the message, its duration, the subjective experience of anonymity, social distance linked to the perception that the virtual world does not affect the offline world, or the transnationality of content (Isasi & Juanatey, 2016).

Authors such as Martín et al. (2021), alluding to other authors, argue that educational research applied to adolescent figures has been an interesting topic from a scientific point of view because there have been important social and emotional changes in this regard, their traits of idealism, egocentrism, or tendency to engage in reckless risk behaviors grow in digital societies. This process is associated with age-specific needs to configure identity and strengthen the feeling of belonging to a group, promoting the abuse of social networks.

2. Objectives

This text is born within the framework of a project for the creation of an online network, a monitoring team, and a telephone app to counteract hate crime tactics with funding from the EU Justice Commission Rights, Equality, and Citizenship Programme (2014-2020). The C.O.N.T.A.C.T. project (Creating Online Network, monitoring Team, and phone App to Counter hate crime Tactics) is a two-year initiative (2015–2017) in which ten countries participate: Cyprus, Greece, United Kingdom, Spain, Italy, Poland, Denmark, Lithuania, Romania, and Malta, through the synergy between 5 Universities and 7 NGOs. Among the objectives pursued by this project are those of improving, promoting, and analyzing the reporting of hate crime and speech, as well as training and raising awareness about what hate crime is, how to prevent it, and how to report it.

The central objective of the research carried out within the framework of the project focused on deepening the perception and knowledge about hate speech present in Spanish youth.

3. Methodology

The research design used is mixed cross-sectional, non-correlational descriptive. The sampling carried out was non-probabilistic and intentional.

Sample

The subjects of the research are young Spaniards from the autonomous communities of Castilla-La Mancha, Extremadura, Madrid, and Catalonia. A stratified probabilistic sample (by age range), made up of $N=186$ (trust=95% and error= ± 2.5) where the number of men ($n=28$; 15%) and the number of women ($n=158$; 85%) varies considerably, with a majority presence of women. On the other hand, in terms of age, the subjects are distributed, in absolute terms and percentages of the sample, as follows: $n=145$ aged between 18 and 23 years (78%); $n=18$ aged between 24 and 29 years (9.7%); $n=23$ aged between 30 and 35 years (12.3%).

For the interviews, a non-probabilistic sample of 20 young people (50% men and 50% women) who had previously participated in the questionnaire phase and who helped delve into the topics discussed around the aforementioned objectives was formed.

Instrument

An ad-hoc questionnaire was designed to find out the opinion of those surveyed about communication on the Internet regarding the perception of hate speech present in Spanish youth. It is a single questionnaire made up of three blocks of questions:

- 1) Regarding public comments taken literally from the internet, they are asked to respond according to a Likert scale of four values: 1=acceptable; 2=acceptable in some way; 3=barely acceptable and 4=unacceptable.
- 2) Related to having ever witnessed threats or insults to another person for reasons of nationality, skin color, ethnic origin, religion, gender, sexual orientation, and others, for which the Likert scale offered to answer consists of four values, being 1=Yes, frequently; 2=Yes, sometimes; 3=No, and 4=the option of Does not know/No answer is offered, which has been understood as a missing value for the purposes of data analysis. Regarding this block, we ask about the space where they have witnessed it (*Work, Educational Centers, Street, Public Transport, Internet, Social Networks*).
- 3) Relating to knowledge about hate speech and hate crimes. In this block, the Likert scale for the answers also consists of four values: 1=disagree; 2=somewhat agree; 3=agree, and 4=strongly agree.

Also, at the beginning of the questionnaire, they were asked for personal data, such as gender and age. After the design of the questionnaire by the research team, it was passed on to the sample of students described above, and, subsequently, the validity and reliability of the three dimensions mentioned were analyzed.

The interviews aimed to allow the young respondents to offer their opinions and talk about their communication experiences (online and offline) and, more specifically, about comments that could be labeled as hate speech openly and flexibly. The interviews were categorized with the same three initial blocks with open questions based on the items prepared for the questionnaire: Public comments taken literally from the internet; Presence of threats or insults for reasons of nationality, skin color, ethnic origin, religion, gender, sexual orientation, and others; Knowledge of hate speech and hate crimes.

Process

Before answering the questionnaires and interviews and obtaining consent for participation in the study, its general-purpose was explained to the participants. The young people were informed that their participation was voluntary and anonymous, respecting at all times Organic Law 15/1999, of December 13th, on data protection.

The questionnaires were completed in approximately 8-10 minutes online using a link provided through a QR code. During the administration of the instrument, at least one researcher was present in the classroom. No participant reported difficulties in answering it.

For its part, the Llere Socio-educational Association conducted 20 virtual interviews in Spain. The interviews were conducted with women and men between 18 and 35 years of age from various towns and

cities in the province of Toledo and Madrid. The duration of the interviews was 15 to 20 minutes and the adopted approach followed the guidelines agreed upon within the framework of the project. The interviews began with an initial presentation by the interviewer on how communication takes place on the Internet and continued with an analysis of hate speech and its implications. The interviews were a second step in the research on the perceptions and impact of hate speech on young people, based on a questionnaire that was prepared, disseminated, and analyzed before the interviews.

4. Discussion/Results

4.1. First and second grouped block: Public spaces and hate speech.

We asked through the questionnaire about the degree of acceptance of comments appearing in public forums through six examples taken literally from the Internet (Likert scale of four values: 1=acceptable; 2=acceptable in some way; 3=barely acceptable; 4 =unacceptable). The examples were taken from the Spanish national context covering 3 topics: migrant/refugee; LGBT; race/minorities, and the examples had a negative message. As we can see in Table 1, the majority tendency in all cases is towards the consideration that all these statements are unacceptable and/or barely acceptable. We are particularly struck by the high deviation (1.29) and the average closest to unacceptable in the case of "*homosexuality is a diabolical evil that must be eliminated*".

Table 1: Public comments and hate speech

	<i>Average</i>	<i>Median</i>	<i>Deviation</i>
<i>"Immigration, especially African immigration, has only brought problems to Spain"</i>	3.79	4.00	0.54
<i>"Islam is cancer and cancer must be rooted out"</i>	3.84	4.00	0.51
<i>"(Immigrants) pay half the taxes of natives"</i>	3.40	4.00	0.83
<i>"Homosexuality is a diabolical evil that must be eliminated"</i>	3.08	4.00	1.29
<i>"As a creature that makes mistakes, I wanted to ask you something, because I'm sure you can fraternize with the fact of being wrong. What if it turns out that homosexuality really is wrong and you are mistaken? It could be."</i>	3.71	4.00	0.68
<i>"The reality is that homosexuality is deviant sexual behavior. I'm very sorry but that's the way things are. The natural thing is heterosexuality, that is, sex between a man and a woman."</i>	3.78	4.00	0.62

Source: Own elaboration

The comments you have read in the previous question were written publicly about articles in online newspapers. To the question posed, do you think you would have marked different options if they had been written in private contexts (e.g., private emails, private WhatsApp chats, private Facebook/Instagram chat messages, etc.)? 87% answered No, 9.1% answered Yes, and 9.1% Does not know/No answer.

If we ask if they have been the object of insults and threats in the first person, the results (Table 2) show relevant data mainly around gender and religion.

Table 2: Object of insults or threats for reasons of hate

	<i>No</i>	<i>Yes, sometimes</i>	<i>Yes, frequently</i>	<i>Does not know/No answer</i>
<i>Nationality</i>	86.6 %	12.4 %	1.1 %	0.0 %
<i>Skin color</i>	88.7 %	9.1 %	0.0 %	2.2 %
<i>Ethnic origin</i>	89.2 %	9.7 %	1.1 %	0.0 %
<i>Religion</i>	78.0 %	21.0 %	0.0 %	1.1 %
<i>Gender</i>	48.4 %	36.0 %	15.6 %	0.0 %
<i>Sexual orientation</i>	89.2 %	5.4 %	3.2 %	2.2 %
<i>Other situations</i>	52.2 %	19.4 %	5.4 %	5.4 %

Source: Own elaboration

In this sense, and when asked if they have been present in any situation where insults or threats have appeared, we find that all the items increase considerably towards *Yes, sometimes* and *Yes, frequently* concerning all the elements asked and especially regarding *sexual orientation* and *gender* as we can see in Table 3.

Table 3: Witness hateful insults or threats to another person

	<i>No</i>	<i>Yes, sometimes</i>	<i>Yes, frequently</i>	<i>Does not know/No answer</i>
<i>Nationality</i>	18.3 %	58.1 %	22.0 %	1.6 %
<i>Skin color</i>	19.9 %	55.4 %	23.1 %	1.6 %
<i>Ethnic origin</i>	28.0 %	51.1 %	21.0 %	0.0 %
<i>Religion</i>	25.3 %	53.8 %	16.7 %	4.3 %
<i>Gender</i>	24.7 %	45.2 %	28.0 %	2.2 %
<i>Sexual orientation</i>	15.6 %	49.5 %	33.9 %	1.1 %
<i>Other situations</i>	25.3 %	33.3 %	12.4 %	29.0 %

Source: Own elaboration

Asked about the places where they have been subjected to insults or threats, more than 68% claim to have been the target of insults or threats for reasons of hate, and of these, more than 70% claim to have been at the same time in "educational centers", "a sports context", "on the street", and/or "public transport", and also "on the internet". Only 11% claim to have suffered it at "work".

The data obtained after the analysis of the survey showed a total agreement regarding the unacceptability of the comments obtained on the Internet. This was corroborated in subsequent interviews. There was some consensus that this type of online comment is unacceptable:

"I think it is not acceptable, because there are many people there, not only because they are migrants who are suffering, and I think there has to be... There has to be equality"

Especially for its scope and intent

"You can express that in a way that doesn't hurt anyone. I don't know, I don't agree with how it's written, at least"

And its lack of rigor:

"It is not positive news and it could be false, but other than that I do not see it as acceptable, because everyone can have their own opinion and express it. That's a lie, but people can express themselves freely."

However, while it is true that the young people interviewed mainly rejected these comments, some framed them under freedom of expression and considered that, even if some of the comments were potentially repugnant, they should still maintain a space within a democratic context:

"I think that is not acceptable, but from what I think I understand concerning freedom of expression, everyone can say what they want, so in this sense, I think it is acceptable because we are free to express whatever we feel or think".

They assumed that these types of comments are made by people who take advantage of the internet as an anonymous safe haven that allows them to extremely express their opinion and that they would not do so in the offline world.

"They don't really show themselves; they're talking through a screen and they write what they think. When they are face to face, if you were to tell a friend on WhatsApp or social media you write one thing, and then if you have to say it to their face, you will not say the same things or in the same way."

Finally, they noted that some of this assessment of acceptability could change if those comments were made in a private context:

"When you are in a private context, maybe you can be more honest";

Although they link this type of comment in a satirical and humorous context:

"Yes, sometimes I realize that I used a word, that I used it with sarcasm and irony, and sometimes I used racism to create humor."

4.2. Third block: Knowledge and position against hate speech.

In this block, we focus on deepening the knowledge that Spanish youth have about the concept of hate speech. In this sense, we ask if they have ever heard of the concept "Hate Speech" and 43% respond that they do not know it. On the other hand, 49% consider that they do know the term and 8% respond that they Don't know /No answer (perhaps we could understand this answer within the Do not know block).

Next, we ask if they think there should be laws against the different ways in which Hate Speech is presented. As we can see in Table 4, in all cases the affirmative answer exceeds 85%. In this sense, it is important to highlight the high percentage of young people who doubt through the Don't know/No answer in all cases.

Table 4: Position on legislating against this type of hate speech

	% Yes	% No	% Does not know/No answer
<i>"Hate Speech" means making negative comments based on prejudice about a person because of their race, nationality, ethnic origin,</i>	87.6	3.8	8.6
<i>"Hate Speech" means insulting a person because of their race, nationality, ethnic origin, religion, sexual orientation, or gender.</i>	89.8	0.0	10.2
<i>"Hate speech" means threatening a person because of their race, nationality, ethnic origin, religion, sexual orientation, or gender.</i>	91.4	1.6	17.0
<i>"Hate Speech" means encouraging other people to show hatred or be violent towards other people because of their race, nationality, ethnic origin, religion, sexual orientation, or gender.</i>	89.8	0.0	9.1

Source: Own elaboration

One of the key questions that appeared in the questionnaire was the presentation of various examples of degrading and discriminatory comments, to obtain the opinion of the respondents on whether it was acceptable for them to appear in a digital context.

After this first part of the interview, respondents were provided with a definition of hate speech. We use the definition of the European Commission against Racism and Intolerance (ECRI) of the Council of Europe, in Recommendation No. R (97) 20 of the Committee of Ministers to the Member States on "Hate speech".

"(...) The term "hate speech" shall be understood to encompass all forms 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 nationalism and ethnocentrism aggressions, discrimination and hostility against minorities, migrants, and persons of immigrant origin".

As well as the approach to the concept of cyber-hate used by the main NGOs that fight against hate crimes and hate speech at an international level.

"Any use of electronic communications of information to disseminate anti-Semitic, racist, intolerant, extremist, or terrorist messages or information"

Regarding this topic, practically all the interviewees were unaware of the concept and definition of "hate speech" and "cyberhate":

"I did not know it"

In any case, after reading the definition, everyone could generally detect that some of the comments collected and displayed from the news could be integrated into those descriptions.

"Do you think this definition refers to what we've been talking about so far?"

- Yes".

Having determined what is meant by hate speech, and knowing the examples mentioned above, they were asked about the possibility of legally regulating this type of communication. For the most part, the interviewees did not know if there is a legal framework that regulates or penalizes hate speech in Spain:

"There are no laws, I can tell you from my own experience"

Only in some specific cases, in which they assumed or were aware of its existence, they could not describe what this legal framework consists of or regulates.

"I think there are hate crime laws, but I'm not sure."

Or if this law is being implemented:

"Yes, I think there are some (laws) but they are not being implemented properly."

In general, all agreed that this type of law should exist, although concepts are confused (hate speech and hate crime):

"Yes, they should exist and I don't know if it should be called a 'hate crime' or just a 'crime' but it should exist."

Furthermore, they thought that the punishment could be a fine to educate the person who uses hate speech.

"Maybe if they were fined, people wouldn't use those kinds of expressions."

Another big area of research, included in the interviews, was related to the question of whether they had suffered in person or had witnessed hate speech against others. Concerning the first question, and once the meaning of hate speech has been clarified, some of the interviewees acknowledge having suffered in the context of bullying in schools:

"Yes, but it was bullying in schools"

"Yes, me and another girl and another disadvantaged boy" (suffered it)

And both online:

"I was bullied with comments on Twitter; they used a nickname and said humiliating things about me"

And offline, that was connected to the online world. But in general, they do not believe they have suffered directly.

"Personally, I don't think so."

Regarding the question of whether they have ever been the authors of hate speech, most of them did not believe that they expressed themselves in that way, or at least not intentionally.

"No, not consciously."

Some of those interviewed have witnessed hate speech used against others.

"Yes, I had a classmate who was Romanian, and he sat at a table, a little apart from the whole class, and everyone insulted him, even the teacher once asked him if his mother didn't iron his clothes"

Hate speech was mostly seen on social media

"I also read comments from friends of my friends who laugh at disadvantaged people or people who have some kind of problem, and I don't like it. In fact, I always tell him. Insulting disadvantaged people or people from another country, mainly those two."

But hate speech also appears on the street

"If you go to buy ice on a Friday at a Chinese store, there are young people who enjoy joking with the owner, imitating his accent when they speak Spanish. – Is it common for things like this to happen? – Yes, very."

In the search for the causes of why these types of comments are made, some of the interviewees did not identify stereotypes and prejudices as motivating elements of aggressive and discriminatory behavior towards different people but rather referred to the low self-esteem of the people who use hate speech and how they use it to compensate for this feeling of inferiority

"Many times, I think it's because of envy or a social situation, I meant, I think it's envy because they insult";

"It is cowardice because it is used to increase self-esteem."

On the other hand, some others found that the bias motivation is behind hate speech, highlighting that some of the motivations are those related to the country of origin, sexual orientation, or physical condition.

"It's based on physical appearance, race, homosexuality, or culture."

"They call her a fat cow, because of the way she dresses."

In terms of style, the interviewees pointed out that hate speech linked to political issues is more aggressive

"For example, when it comes to politics, there are super aggressive people and yes, I indeed read a comment where people try to give their opinion calmly, but that is 1% because the rest of the comments are all insults and verbal attacks "

On the hate speech linked to other prejudiced motivations, it is hidden under the excuse of humor or dark humor style.

"The joke was not intended to hurt anyone."

Finally, concerning the alternatives and possibilities of denouncing hate speech when it goes beyond the limits of what is acceptable or it is proven that it is part of an aggressive strategy, most of the interviewees did not know how to denounce it:

"If I see a comment, I can go to the police station and tell them I saw it and want to report it, but I'm not sure how this works."

Some know how to report a comment on the social platform itself, but mostly they don't know how to do it legally:

"Not much, not much really. How can I report that? If it's through social media, there's a way you can. If it's offensive content, you can report the profile to Facebook and ask for it to be removed or if they can take it down, so you can report it, let's say socially, on social media, of course".

Regarding the question about if they knew the alternatives to report, what would they do, mainly they would look for an alternative instead of reporting:

"No, I do not report. I will try to solve it myself "

"I don't know. It depends on the situation and also if I'm involved, because if I see that on social media and I don't know the person, I can't report anything. Maybe because I don't have the information on how to do it, if we knew, maybe even anonymously, like, hey, look at this site, there's a comment on that situation, yes, it could be ".

5. Conclusions

In general terms, young people consider that hate speech is part of communication on the internet and social networks. Although indeed, they did not identify it on the street with the same presence, and its scope and impact are considered different from that of the online world, they understand that both feed off each other. They understand that its dissemination is facilitated more easily by the theoretical anonymity provided by social networks. In the network, insult is the main manifestation of harassment on the internet (Romera et. Al., 2017).

They understand that hate speech is used against social groups determined by their physical, cultural, ethnic, and sexual orientation, and although they may come to recognize that in the private sphere they have used it in a humorous tone, they are not aware of having suffered this, though they have seen it. They also identify hate speech as an intimidation strategy. Our research converges with previous studies such as Al Serhan et al. (2020) whose results confirmed that the majority of respondents were able to identify hate speech on social media sites (77%) and stated that they were exposed to hate speech when using the Internet (83%), especially through Facebook and Twitter.

Initially, participants had generally heard of it but did not know what hate speech is, if and when it is punishable, and whether there are laws that allow it to be prosecuted. They also don't know much about spaces and tools to report hate speech, beyond those available on platforms. In general, they would look for alternatives to deal with it before the police report or the legal complaint, which they consider has little impact. Therefore, hate violence is considered to be the result of hate incidents and crimes that negatively impact the physical and mental health of victims in the short and long term (Gil-Borrelli et al., 2018; Gil- Borrelli et al. al., 2020).

Animosity and hatred towards certain groups that support hate speech and its reproduction online, cyber-hate, are also reflected in other forms of expression of violence and intolerance, such as cyberbullying. Based on the study carried out, the young people interviewed, in their real and virtual lives, experience, face, and observe as witnesses how the expression of hate occurs widely on the Internet and identify that in their personal experiences this expression of hate is closely linked to bullying and serves as a justification for exercising it. The identity of the people is presented, as well as a throwable element to exercise hate speech. This research presents limitations to determine differences by sex and age that could yield relevant information that allows us to design proposals aimed at raising awareness among the public in general and young people in particular.

In this context, large amounts of hate comments are published daily which have led to the development and promotion of projects that can contribute to mitigating these effects such as the PHARM project which aims to monitor and model hate speech against refugees and migrants in Greece, Italy, and Spain (Vrysis et al., 2021). For its part, the C.O.N.T.A.C.T. project (Assimakopoulos et al., 2017) focuses on promoting general preventive measures, such as the collection and scientific analysis of data that help to better understand the context of online hate speech, as well as the development of training sessions aimed at stakeholders (young police and media) to build a stronger civil society. It also offers reactive measures and response mechanisms, such as the creation of a dedicated web platform and phone app for reporting hate incidents. Or the SHELTER Project (Support and advice through the health system for hate crimes victims), which places health services at the epicenter of reporting hate crimes and supporting and accompanying their victims to fight against racism, xenophobia, and any other form of intolerance and whose objective is focused on increasing the capacity of health services and, in particular, of their personnel, such as doctors and nurses, to identify and advise victims on their rights (Moreno and Arroyo, 2021). Further analysis is needed of how user practices and platform policies relate to racism and shape contemporary racism (Matamoros-Fernández & Farkas, 2021).

In short, the fight against hate on the Internet is a multifaceted issue, where legal measures must be accompanied by proposals and responses of a social and educational nature, and must have the institutional commitment and that of the companies that provide technological services, all aimed at building a safer and more respectful online space for all, in which far from giving more fuel to those who spread messages of hate hidden by the anonymity offered by the online medium, the visibility of those narratives that contribute to promoting truly democratic values respectful of diversity is promoted.

6. References

- Anti-Defamation League (2010). *Responding to Cyberhate, Toolkit for Action*. Anti-Defamation League. <https://www.adl.org/sites/default/files/documents/assets/pdf/combating-hate/ADL-Responding-to-Cyberhate-Toolkit.pdf>
- Al Serhan, F., & Elareshi, M. (2020). New media and hate speech: A study of university students in Jordan. *Opción: Revista de Ciencias Humanas y Sociales*, (26), 166-184.
- Assimakopoulos, S., Baider, F. H., & Millar, S. (2017). *Online hate speech in the European Union: a discourse-analytic perspective*. Springer Nature.
- Ayuntamiento de Barcelona (2017). *Informe Discurso del Odio. Resumen Ejecutivo*. Ayuntamiento de Barcelona. http://ajuntament.barcelona.cat/bcnvsodi/wp-content/uploads/2017/02/Informe_Discurso-del-Odio_resumen-ejecutivo_ES-2.pdf

- Ben-David, A., & Fernández, A. M. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *International Journal of Communication*, 10, 27.
- Burnap, P., & Williams, M. L. (2014). *Hate speech, machine classification and statistical modelling of information flows on Twitter: Interpretation and communication for policy decision making*. Oxford U.K.
- Cabo, A., Tarragona, L., & Vallès, O. (2015). *Informe del Observatorio PROXI* (Vol. I.). Observatorio PROXI. <https://www.observatorioproxi.org/images/pdfs/INFORME-proxi-2015.pdf>
- Cabo Isasi, A. y Juanatey, A. (2016). *El discurso de odio en las redes sociales: un estado de la cuestión*. Ayuntamiento de Barcelona. http://ajuntament.barcelona.cat/bcnvsodi/wp-content/uploads/2017/02/Informe_Discurso-del-Odio_resumen-ejectivo_ES_2.pdf
- Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 512-515.
- De Haro, J. (2019). *Redes sociales en Educación*. http://eduredes.antonio Garrido.es/uploads/6/3/1/1/6311693/redes_sociales_educacion.pdf
- De Salvador Carrasco, L. (2012). *Redes de anonimización en Internet: cómo funcionan y cuáles son sus límites*. Instituto Español de Estudios Estratégicos (IEEE).
- Del Río, J., Sádaba, C. y Bringué, X. (2010). *Menores y redes ¿sociales?: de la amistad al cyberbullying*. <http://www.injuve.es/sites/default/files/RJ88-09.pdf>
- Ekman, M. (2019). Anti-immigration and racist discourse in social media. *European Journal of Communication*, 34(6), 606-618. https://doi.org/10.1177/0267323119886151_journals.sagepub.com/home/ejc
- European Commission (2016). *Code of conduct on countering illegal hate speech online*. http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf
- Farkas, J., Jannick S., & Neumayer, Ch. 2018. Cloaked Facebook Pages: Exploring Fake Islamist Propaganda in Social Media. *New Media & Society* 20(5), 1850-67.
- Gil-Borrelli, C. C., Martín-Ríos, M. D. y Rodríguez-Arenas, M. Á. (2018). Propuesta de actuación para la detección y la atención a víctimas de violencia de odio para profesionales de la salud. *Med. clín*, 150(4), 155-159.
- Gil-Borrelli, C.C., Dolores Martín-Ríos, M. D., López Corcuera, L., Reche Martínez, B., Torres Santos-Olmo, R., Muriel Pati, E. y Rodríguez-Arenas, M. A. (2020). Elaboración de un cuestionario de detección de casos de violencia de odio en urgencias hospitalarias. *Gac Sanit* 34(2). <https://doi.org/10.1016/j.gaceta.2019.01.006>
- Holcomb, L., & Beal, C. (2010). Capitalizing on web 2.0 in the social studies context. *TECHTRENDS*, 54(4), 28-32. <https://dx.doi.org/10.1007/s11528-010-0417-0>
-

- Instituto Nacional de Estadística (2021). *Encuesta sobre equipamiento y Uso de Tecnológicas de Información y Comunicación en los Hogares 2021*.
- Isasi, A. C. y Juanatey, A. G. (2016). *El discurso del odio en las redes sociales: un estado de la cuestión*. https://plataformaciudadanacontralaislamofobia.files.wordpress.com/2017/03/informe_discurso-del-odio_es.pdf
- Jefatura del Estado. (1999, 13 de diciembre). *Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal*. Boletín Oficial del Estado. <https://www.boe.es/buscar/pdf/1999/BOE-A-1999-23750-consolidado.pdf>
- Lamerichs, N., Dennis, M. C., Puerta, R. R., & Lange-Bohmer, A. (2018). Elite Male Bodies: The Circulation of Alt-Right Memes and the Framing of Politicians on Social Media. *Participation, 15*(1), 180-206.
- López-Berlanga, C. y Sánchez-Romero, C. (2019). La interacción y convivencia digital de los estudiantes en las redes sociales. *Revista nacional e internacional de educación inclusiva, 12*(2), 114-130.
- Martín-Martín, M., Asensio-Muñoz, I., y Bueno-Álvarez, J. (2021). Uso de las redes sociales en estudiantes de secundaria: análisis de perfiles para la intervención educativa. *REVISTA COMPLUTENSE DE EDUCACIÓN, 32*(3), 303-314. <http://dx.doi.org/10.5209/rced.57189>
- Matamoros-Fernández, A., & Farkas, J. (2021). Racism, hate speech, and social media: A systematic review and critique. *Television & New Media, 22*(2), 205-224.
- Ministerio del Interior (2020). *Informe sobre la evolución de los delitos de odio en España*. <http://www.interior.gob.es/documents/642012/13622471/Informe+sobre+la+evoluci%C3%B3n+de+delitos+de+odio+en+Espa%C3%B1a+a%C3%B1o+2020.pdf/bc4738d2-ebe6-434f-9516-5d511a894cb9>
- Moreno-López, R., & Arroyo-López, C. (2021). *Support and advice through health system for hate crimes victims: a socio-sanitary approach*. Tirant Lo Blanch. <https://editorial.tirant.com/es/libro/support-and-advice-through-health-system-for-hate-crimes-victims-a-socio-sanitary-approach-arroyo-lopez-cesar-9788418656842>
- Olweus, D. (1999). *Conductas de acoso y amenaza entre escolares*. Morata
- Olteanu, A., Castillo, C., Boy, J., & Varshney, K. (2018). The effect of extremist violence on hateful speech online. *Proceedings of the international AAAI conference on web and social media, 12*(1).
- 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>
- Pohjonen, M., & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. *International Journal of Communication, 11*, 19.

Romera, E. M., Ortega, R., Rey, R. D., Casas, J. A., Viejo, C. M., Gómez, O. y Luque, R. (2017). *Bullying, cyberbullying y dating violence: Estudio de la gestión de la vida social en estudiantes de Primaria y Secundaria de Andalucía*. Fundación Pública Andaluza, Centro de Estudios Andaluces.

Sacristán, A. (2013). *Sociedad del conocimiento, tecnología y educación*. Ediciones Morata.

Save the Children España (2016). *Yo a eso no juego. Bullying y cyberbullying en la infancia*. https://www.savethechildren.es/sites/default/files/imce/docs/yo_a_eso_no_juego.pdf

Vrysis, L., Vryzas, N., Kotsakis, R., Saridou, T., Matsiola, M., Veglis, A., Arcila-Calderón, C., & Dimoulas, C. A. (2021). Web Interface for Analyzing Hate Speech. *Future Internet*, 13(80). <https://doi.org/10.3390/fi13030080>

AUTORES:

Roberto Moreno López

Diploma in Social Education in 2003 and Doctor in Humanities and Education in 2017 from the Universidad de Castilla-La Mancha. He has developed his research and teaching work in the area of Pedagogy for a decade as a member of the GIES-CLM, TRES, and SIMI groups at the Universidad de Castilla-La Mancha and the Universidad Internacional de la Rioja. His main lines of work focus on diversity, educational and social inclusion, and hate speech and crimes.

roberto.moreno@uclm.es

Índice H: 4

Orcid ID: <https://orcid.org/0000-0002-6238-9440>

Google Scholar: <https://scholar.google.es/citations?user=SU3WbRoAAAAJ&hl=es>

ResearchGate: <https://www.researchgate.net/profile/Roberto-Moreno-8>

Scopus ID: 57195273852

Cesar Arroyo López

Doctorate in Humanities, Arts, and Education from the Universidad de Castilla – La Mancha. Bachelor of Humanities. Specialist in management, planning, and social innovation. Expert in socio-educational intervention with families and children at risk. Specialist in international development cooperation. Expert in European Programs. He has worked in the social field for more than 15 years, the last 8 of which he participated in different initiatives at a European level.

cesar.arroyo@ineeys.es

Orcid ID: <https://orcid.org/0000-0002-8249-5697>