Master of Arts in Dispute Resolution

RESTORATIVE PRACTICE AS AN ALTERNATIVE TO ADDRESS
HATE SPEECH ON SOCIAL MEDIA

By

Mariane Nishi Gomes kobori

51712610

May ─ 2022

Dissertation Submission Form

LEARNER NUMBER:	51712610
NAME:	Mariane Nishi Gomes Kobori
COURSE:	MA in Dispute Resolution
DISSERTATION TITLE:	Restorative Practice as an alternative to address hate speech on social media
SUPERVISOR NAME:	Gareth Leech
WORD COUNT:	15.665
DUE DATE:	20/05/2022
DATE SUBMITTED:	20/05/22

I certify that:

- This dissertation is all my own work, and no part of this dissertation has been copied from another source: Yes √ No ☐
- I understand that this dissertation may be checked for plagiarism by electronic or other means and may be transferred and stored in a database for the purposes of data-matching to help detect plagiarism: Yes √ No ☐
- This dissertation has not previously been submitted in part or in full for assessment to any other institution: Yes √ No ☐
- I consent to the inclusion of my dissertation in an electronic database maintained by the library: Yes √ No ☐

(Note that if no check boxes are selected, this dissertation will be regarded as NOT properly submitted, and may be subject to a late submission penalty)

Name: Mariane Nishi Gomes Kobori

Date: 20/05/2022

Notes:

SUMMARY

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I want to thank you, my family, without them I would never have arrived so far. I want to thank my friends that always have supported me and always have heard my complaints when things were not that nice. I also wanted to thank all the teachers that have shared their knowledge with us, and in special to my supervisor Gareth Leech, if it wasn't for him, I wouldn't be as interested in this subject as I am now and has oriented and supported my ideas since the first day.

A very special thank you to the group Drumcondra Power, a place where I have shared anguish and yarning and always have found support and a good laugh. I wanted to thank every single person that has been with me during this journey, however, would not have pages enough in this work to mention all of them. If you ever heard and seen my complaints, my tears, my celebrations I just want to say thank you because you were a part of this too.

ABSTRACT

People's behaviours and relationships are altering as a result of technological advancements, the Internet, and social media. It is now simpler to locate an online community of nearly anything because of the ability to assemble people who have many things in common, particularly those who have a similar worldview. The number of users has grown over time, as has the number of abuses and crimes perpetrated inside the social media sphere.

It is essential to analyse the effects of hate speech disseminated on social media these days, as well as comprehend the reasons and current measures in place to try to address the issue. The background and development of hate speech on social media will be objectively assessed in this work for the reader to achieve a better understanding of the issue and analyse the efficiency of the restorative practice applied to address hate speech on social media. This work will be evaluated by the appreciation of reliable research and survey made with social media users to produce accrued research.

INTRODUCTION

Hate speech is an objectionable manner of speaking that employs stereotypes to transmit a hate ideology intended at someone based on their traits, such as their gender, ethnicity, religion, race, condition, or even political viewpoints. Nevertheless, because people have various parameters and points of view, some people may regard a given comment to be hate speech, while others may claim that it is genuinely free speech. Policies aimed at reducing hate speech risk and constraining free expression are inconsistently implemented.

Social media, like the internet, was based on the ideas of open communication and free information flow. Online communications are used to abuse, harass, harm, and intimidate people, thereby smothering their opinions and ideas. As per recent research, there is a higher prevalence of aggression when there is a large amount of hate communicated on social media. Social media is a product of technical advancement, and its appeal among young people has grown rapidly. It has the ability to bring people together and build a solid network. Simultaneously, the downside is that it has led to an increase in hate speech, which is one of the most popular and widespread global problems.

Restorative practice is an approach that could be used as a tool to address this problem. It has been proven by studies the effectiveness of restorative practice on the physical grounds. Restorative justice is a form of strategy that focuses on the harm, prioritises the needs of victims, encourages offenders to accept responsibility for their actions, recognises the grief they've caused, and, if certain conditions are met, gives a path to redemption. An important notion is that those most affected by the injury should be able to actively engage in its rehabilitation. The internet was created on the ideas of open communication and unfettered information flow. In reality, internet technologies are being used to harass, hurt, and ridicule individuals, thereby smothering their voices and ideas. (Salehi, 2020)

This research is particularly significant in assessing the influence of hate speech transmitted on social media and understanding what techniques platforms are employing to address the issue and whether such efforts are successful. Additionally, the research will be useful in determining how successful an alternate solution may be in addressing the issue. Finally, the study will look at alternative solutions and successful strategies to cope with a worldwide problem. As a result, the debate and pursuit of alternatives are still important and relevant.

1. LITERATURE REVIEW

Introduction

In this chapter, the subject will cover fundamental concepts, terminology, history, and background. It will provide the reader with a good understanding of the subject. The issue will be analysed and synthesised. It will be able to address how the problem arises and spreads, and it will be able to identify how the subjects are related using. It will go through the present methods in place to manage and fix the problem, as well as how a new strategy can be advantageous and successful in tackling it.

1.1. Hate Speech, Online Harm and Free Speech Overview

1.1.1. Hate Speech, definition and how it develops.

The concept of hate speech can be easy to understand for most people. However, it is not often that people think deeply about its concept and realize the dimension that its effects could cause. Nowadays, it has become a widespread concept due to its popularity on social media.

Hate speech is more than a label for a group of expressions. It can also be used as a negative evaluative phrase or as a subject for censure. As a result, defining this category has significant consequences. (Philosophy, 2022)

According to the definition of the No Hate Speech Movement Ireland, hate speech includes any communication that disseminates, provokes, encourages, or seeks to justify any sort of hatred, stereotyping or discrimination based on intolerance. (Ireland, n.d.) This type of harmful expression and communication can be present in different types and areas such as gender, race, beliefs, and others. Due to the numerous variations, this work will briefly explain only five primary types: sexism, racism, religious intolerance, political intolerance, and homophobia/ transphobia/ LGBTIQ+ intolerance.

Sexism is a sort of hate speech that is entirely motivated by gender. All expressions that disseminate, provoke, or excuse hatred based on sex are considered sexist hate speech. This type of hate speech has a higher incidence among women for various reasons.

According to the Council of Europe Gender Equality Strategy (2016), Sexist hate speech is influenced by several factors, such as the prevalence of patriarchal societies, the spread of

derogatory communications about women or girls, and violent and sexualized images, particularly in the media, and societal expectations about women and men's sexuality and roles. In addition to that, The fact that many targeted women do not report sexist hate speech hides the actual magnitude.

Racism is a manifestation of hate speech based on the belief that distinct groups of individuals who have different hereditary behaviours, qualities, and cultures may be divided based on the superiority of one race over another. According to No Hate Speech Movement Ireland, racism occurs "when an individual, structure or institution intentionally or unintentionally abuses their power to the detriment of people, because of their actual or perceived "racialised" background. Therefore, racism is more than just prejudice but the combination of power and prejudice." (Ireland, n.d.)

Religiously motivated hatred and prejudice are generally distinctive to the targeted group and are moulded by stereotypes, myths, and misconceptions about various faith and belief groups. (Ireland, n.d.) In this context, religious intolerance is similar to racism as they both can be based on the superiority of one belief over another.

In the case of political intolerance, the hate is more towards an intolerance related to an ideology and perhaps interests. Some authors state that this is commonly characterized as the extent to which we grant political rights and civil liberties to groups or persons we disagree with. (Marcus, Sullivan, Theiss-Morse, & Wood, 1995). People generally choose their political positions, focusing on how each political party can attend better to their demands and which political party's beliefs are most close to theirs. Political intolerance might arise when people realise that their interests will not be defended. Discussions around the subject are deep and suggest that it is necessary to talk about tolerance in a sense that until where the other party will tolerate not having their interests satisfied.

In addition to the definitions, it also needed to pay attention to how hate speech develops and at what point the borders of hate speech go beyond the online environment and start to become a threat in real life. In this matter, The Anti-defamation League has developed a pyramid to help better comprehend the distinctions between biases and hatred-based behaviours.

The Pyramid depicts the pervasiveness of prejudice, hatred, and injustice in civilization. It is structured in ascending layers of attitudes and conduct that increase complexity from the bottom to the top. The top levels, such as a pyramid, are maintained by the lower ones, but,

unlike a pyramid, the levels are not created sequentially or illustrate each level's hierarchy. Bias at every level represents an oppressive system that harms individuals, institutions, and society. Unchecked bias can become "normalised," contributing to a societal habit of condoning prejudice, violence, and inequality. (league, 2021)

Pyramid of Hate © 2021 Anti-Defamation League
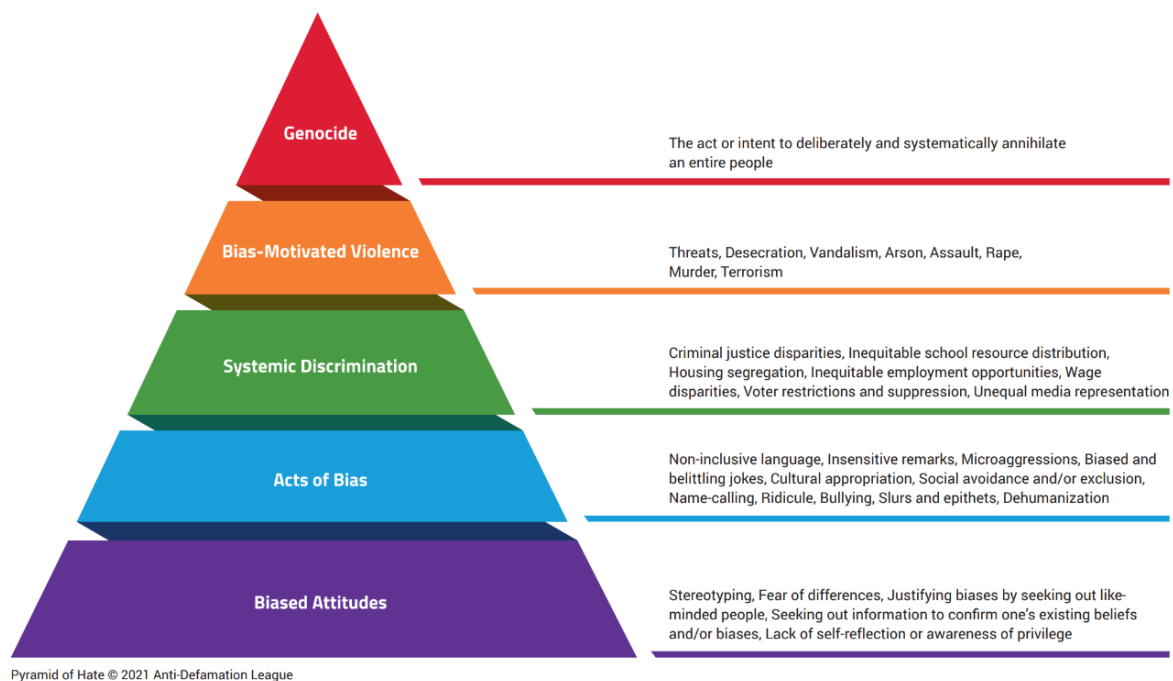
(Source: Anti-Defamation League, 2021)

The pyramid's base is built on biased attitudes. This stage is when the person has a pre-concept that generates a preference in favour or against a specific individual or group, a preference that interferes with fair judgment. The pyramid above shows examples of those biases, such as stereotyping, fear of difference, and others. It is essential to consider that everybody naturally has a bias. However, the focus here is when this bias is reflected in our behaviour towards others.

The second level is the acts of bias. This level is frequently and quickly found as hate speech on social media. It is when the bias becomes attitudes and is externalised through actions such as bullying, microaggressions, name-calling, etcetera. Like the previous one, this level is equally destructive to the individual targeted by the behaviour, but they can also express notions that demeanour or devalue the members of identity groups. Furthermore, the actions can be intentional or unintentional, but they are harmful and detrimental.

The third level is systematic discrimination. This level is defined as the unjust treatment of one individual or a group because of some feature of that person's or group's identity, such as race, gender, or culture. Discrimination is systemic, which means it is anchored in society, history, and institutions and causes extensive. suffering. Systemic prejudice contributes to bias in the community. The two preceding aspects lay the groundwork for discrimination by reinforcing the target group's lack of earning fair treatment.

Bias-motivated violence is the next stage. This layer comprises occurrences such as vandalism, damage, threats, assaults, and even murder. Individuals or organisations might carry it out in a planned or unplanned manner. However, while the violence may be directed at a single person or a small number of individuals, it can also hurt members of the same identification group. It typically occurs in a culture where people are dehumanised and targeted based on their identity; in this situation, the target persons and the community require assistance.

The genocide is the pyramid's top and final tier. It is the deliberate killing of entire people or communities based on their identities. The most well-known example is the Holocaust, in which the Nazis claimed that one race was superior to others and subsequently perpetrated genocide against Jews, homosexual people, and others.

The pyramid of hate illustrates how bias and hatred are built up layer by layer and how, despite their differences, they all work together to hurt target groups via prejudice, discrimination, and aggression. Even though certain aspects of the pyramid, such as the top two layers the biased attitudes and actions of bias, appear to be a regular part of life, all levels contribute to the mistreatment and abuse of individuals based on their identities, since the lower levels are also used to support the upper levels.

1.1.2. Online Harm Definition

To start understanding and talking about online harm, understanding the concept of harm first is needed. According to Merriam-Webster dictionary (n.d.), harm as a *verb* is defined as "to damage or injure physically or mentally." Before the internet, to be more precise, it was common to think that people could be or would be more likely to be harmed face-to-face in specific environments. As a result of the internet's growth, things have changed, and now people can witness and suffer from harmful actions in the comfort of their own homes.

Harmful content is constantly widespread online, and platforms may be used as tools for bullying, hate, and abuse where it is possible to find much harmful information.

The L&RS Note (2020, online) states that online harm is a word that is becoming more prevalent to refer to media and online communications that have the potential for harm. What is essential to bring to this discussion in this definition is that although this communication has the potential to be harmful to someone, that does not mean that this communication is illegal. According to Gov.Uk (2021), Online harms are not necessarily unlawful, but even legal ones can be detrimental to a person. This fact also is vital to realise the obstacles to coping with hate speech once some harmful content is not classified as illegal. Questions such as how to penalise something that is not considered illegal?

Another point that it is necessary to consider is the limits of what is harmful. The definition and limits of what is harmful or not are subjective and may vary from one person to another. However, the limits and concepts might be changeable and unstable research made by Microsoft has identified what age group suffers the most from online risks. According to the research, compared to other age groups, people between the ages of 18 and 34, also known as millennials, are exposed to the highest levels of online risk and experience the most severe effects. Furthermore, once millennials have had a terrible online experience, they lose trust in people both online and offline; they get agitated, sad, lose sleep, or lose a friend; and they worry that the unpleasant event may happen again. (Beauchere, 2018)

However, as seen previously, government and platforms face a dilemma due to the lack of laws, regulations, and the abstract matter of the subject. Currently, government and social media platforms have been looking to deal with the problem by implementing numerous contingency measures and updating and creating new usage policies. According to L&RS Note (2020, online), many modifications have been proposed in Ireland to address potential gaps and shortcomings in the present legal system. The note also brings The Law Reform Commission (LRC) issued a Report on Harmful Communications and Digital Safety made in 2016 that advised caution when utilising criminal law to address online harm. The paper proposed a hierarchical approach to dealing with the issue that includes:

• Education: to create user empowerment and foster safe and positive digital citizenship.

• Civil law and regulatory oversight: where education and related responses are ineffective, civil law should be favoured as it is less onerous than criminal law.

• Criminal law: only the most serious harm should be subject to criminal law.

In conclusion, even though the government and platforms' effort to cope with the issue still is a grey area that divides opinions, it seems a problem far from being solved and a discussion far from being ended. Everybody wants to keep their right to free speech and be able to express their opinion and stands for their values and beliefs. However, when does this right start harming and segregating other people's rights?

1.1.3. Hate Speech vs Free Speech

The discussion that divides the online world and brings loads of question marks for social media users is whether some content should be censored because somehow it might be harmful to someone, or the content should not be censored because it represents the right of someone to express their opinions. Salehi (2020) states that The internet was founded on the principles of open communication and the free movement of information. However, internet technologies are used as weapons to harass, harm, and belittle others, silencing voices and opinions.

Although it needed to observe that free speech is not a bad thing, it is through free speech that we can access a free press and be able to express our satisfaction or dissatisfaction with the government. The people's opinion is essential. Everybody has the right to express themselves, exchange knowledge, and advocate for a better society. Everybody has the right to agree or disagree with people in authority. Exercising these rights without fear or interference is essential to living in a free and equal society where individuals may obtain justice and fulfil their human rights.

According to Laub (2019), internet hate speech has been linked to an increase in violence and harassment worldwide, including mass shootings, mass murders, and cultural genocide. Nations such as the United States provide social media companies broad control over their information and the execution of hate speech legislation. Nowadays, Platforms have been transformed into manipulation tools able to control and change the course of elections and important decisions.

The most recent example that raised loads of questing, curiosity and worries about the power of manipulation of social media is the fact that Elon musk became the major stakeholder of the platform Twitter due to an eminent interest in changing the platform's

usage policies since he declared that does not agree with the way that the platform measures to contain certain types of content. According to Kleinman, the last thing needed now is a platform that deliberately ignores violent and abusive comments directed at users, especially those who are predominantly harmed, such as women, non-binary people, and many others. (Kleinman, 2022)

The worries about the freedom of speech in social media are not new according to Alter (2022)

"Freedom of speech" has become a paramount concern of the techno-moral universe. The issue has anchored nearly every digital media debate for the last two years, from the dustup over Joe Rogan at Spotify to vaccine misinformation on Facebook. Meta founder Mark Zuckerberg gave a major speech at Georgetown in 2019 about the importance of "free expression" and has consistently relied on the theme when explaining why Facebook has struggled to curb disinformation on the platform.

The social media platforms must have the responsibility to deal with hate speech; however, it is challenging once the control over what people express is needed to do with cautious to not censored too much and transform the platform into a non-democratic space where different opinions are not acceptable. Furthermore, banishing people and content deliberately might not be efficient in coping with the roots of the problem and consequently make people feel excluded and move the problem to another platform. According to the idea of 'out of sight, out of mind,' repression, such as eliminating material, only works against the symptoms. It is consequently critical to offer knowledge regarding hatred's broad patterns, protagonists, and goals. This cannot prevent hatred immediately, but it can limit its harmful effects and spread. (Johannes Baldauf, 2019).

To summarise, hate speech is challenging and subjective. Individuals have distinct parameters and perspectives; some may perceive a comment as hate speech, while others claim it is genuinely free speech. Policies that prevent hate speech risk might constrain freedom of expression that are enforced inconsistently.

1.2. How information spreads on social media and copings mechanisms

1.2.1. The impact of social media algorithms on the transmission of content.

There is no way to talk about social media and not mention how the information gets to the people and how it spreads and becomes famous. According to the dictionary, one factor contributing to a post or information on social media becomes what they call "viral", which means "quickly and widely spread or popularized especially employing social media." (Merriam-Webster, n.d.) is a technology called an algorithm. Golino (2022) states that Algorithms in social media platforms are technological techniques of categorising posts based on relevance rather than publication time to define which information a public sees first based on the expectation that they would actively engage with such information.

Matelski (2018) states that a social media algorithm is a statistical computation used by websites such as Facebook, Instagram, Twitter, YouTube, LinkedIn, and Snap Chat to assess user patterns and personalise their social experience. As a result of people's reactions, the algorithm understands and filters what would be more interesting for people, so consequently, it works to deliver the same type of content all the time. Behind the scenes, social media algorithms optimise the user experience without the user being aware that the system is in action.

Social media started using this selective technology to tackle hate speech using the algorithm system. Once using the system, it is possible to reach an enormous amount of hate content automatically than using people analysing each data separately. According to Time, the company Facebook eliminated more than seven million instances of hate speech in the third quarter of 2019, a 59 per cent rise over the previous quarter. They stated that more and more of that hate speech (80 per cent) is now being caught automatically by artificial intelligence rather than by people. (Perrigo, 2019). However, the system has flaws and might fail to detect the content or put down some random content unrelated to the subject. The same article states the problem that the platform has with the system.

The system is more efficient in some specific languages than in others ones regarding languages. Still, according to Perrigo (20109), The algorithms that Facebook now employs to filter hate speech only function in a limited number of languages. That indicates it is now simpler for Facebook to control the propagation of racial or religious hatred on the internet, particularly in industrialised nations and communities where worldwide languages such as English, Spanish, and Mandarin predominate.

In the same direction, Golino (2022) Algorithms that utilise shadow bans may create information gaps by hiding or ignoring some postings while favouring revenue-generating

ones. This component of algorithmic design is contentious since it implies determining which material consumers should deem significant or worthy of praise. Consequently, may result in a politicised and non-objective choice about who and what receives the focus. As a result, algorithmic design inexorably influences the spread of culture and shapes the digital society in a specific way: it determines which type of content or topic should be prioritised in each content that deserves more visibility than others.

Ultimately, it is essential to point out that even though some answers to the questionnaire for this research show that some people might have never witnessed hate speech, that does not mean that it does not exist or is not there on social media used most. That might mean just social media algorithm is already trying to combat and avoid that certain content gets to them or even that, according to their engagement on the platform is not likely that they come across this type of content.

1.2.2. Content moderation as a contingency measure

One of the tools that social media can use to try to deal with issues is content moderation. Content moderation is a way that platforms can enforce their rule and policies. It is a popular technique used by online platforms that rely largely on user-generated content, such as social networking platforms, online sellers, communities, forums and others. When a user submits material to a website, that content is subjected to a screening procedure (the moderating process) to ensure that it complies with the website's policies, is not unlawful, improper, abusive, and so on. (Basedo, 2020)

Content moderation can be either human or automated such as algorithms systems. Because online communication occurs on a "large scale," human censors are unable to evaluate all information before it is made visible. The vast volume of internet material makes screening, even reported stuff, a challenging endeavour. As a result, social media companies are increasingly relying on AI in the form of automated algorithms that deal with harmful content, including hate speech, proactively or reactively. Technologies for dealing with content such as hate speech are still in their "infancy." The algorithms used to achieve this automation are often tailored to the kind of content, such as images, texts, videos and others. (Alkiviadou, 2021)

However, using automated content moderation using automatic systems and artificial intelligence seems a good idea for dealing with the high volume of content, but this is not always this method is reliable, the system has flaws that can seriously rely on freedom of speech. According to Llansó et al. (2020), the employment of these technologies raises serious problems about the impact of AI on our information technology environment and, eventually, on rights to free speech and the accessibility of information. In addition to that, more advanced AI cannot solve the problems by itself. Tackling hate speech with AI without human monitoring and doing so strategically rather than responsively puts free speech in jeopardy. Simultaneously, the incapacity of technologies to detect the complexities of human interactions, as well as the biases which have shaped the design and operation of such systems, raises difficulties about the concept of non-discrimination. (Alkiviadou, 2021).

To summarize even though social media is constantly trying to develop modern systems and find more effective ways to tackle the issue they encounter challenges such as not having humans enough to analyse all the demands of millions of content thrown at the platform every day by millions of people or having technological systems that work with limitation as it does not understand the subjectivities and complexities of human interactions. Furthermore, this subject raises the question that if bashing people and excluding them is the right approach to try to cope with the issue, once this can be only a remedy and a temporary measure that is not able to reach the roots of the problems.

1.3. Restorative Practice Overview

1.3.1. Restorative Practice definition.

Restorative practice is an approach to addressing various issues related to violence whether physical or psychological. According to PDST (n.d.), restorative practice is founded largely on a set of basic ideals and the intentional cultivation and strengthening of certain skills such as empathy and problem-solving abilities. This approach enables the development of trust between and among individuals. The differential of this approach is that in this model the focus is on the harmed part rather than on the part that causes the harm which is more common for traditional systems of punishment.

In this context, some authors argue that nomenclatures are important when dealing with and applying this approach. According to Downes et al. (2007), one of the goals is to avoid using

phrases like victim and offender and instead refer to the people involved in a dispute. In this case, it could be seen to be beneficial to highlight data as the participants view it. Seeking to illustrate the importance of the acts for the parties involved in the process. The meaning of these acts, how they were meant, as well as how they ended out, is frequently debated, rather than the specific acts themselves. The objective at the foundation of a mediation process is to bring changes in these meanings into the light, so opening doors for changes in the meaning assigned to the acts between one or both sides. In the same line, Katherine & Lorenn (2013) avoid labelling people as victims or offenders, which are based on their experiences and actions. Labels classify individuals and can affect our opinions of them. People are greater than what has happened to them or what they have previously done.

Zehr (2015) refers to these two systems as retributive and restorative justice. In his view of retributive justice, a crime is a state infraction defined by illegal behaviour and culpability. Justice assigns blame and delivers punishment in a dispute between the criminal and the state governed by a set of rules. On the other hand in restorative justice crime is an offence against individuals and relationships. It provides responsibilities to correct the situation. The victim, the offender, and the community are all involved in the pursuit of solutions that encourage repair, healing, and trust.

Retributive Lens	Restorative Lens
Crime defined by violation of rules (i.e., broken rules)	Crime defined by harm to people and relationships (i.e., broken relationships)
Harms defined abstractly	Harms defined concretely
Crime seen as categorically different from other harms	Crime recognized as related to other harms and conflicts
State as victim	People and relationships as victims
State and offender seen as primary parties	Victim and offender seen as primary parties
Victims' needs and rights ignored	Victims' needs and rights central
Interpersonal dimensions irrelevant	Interpersonal dimensions central
Conflictual nature of crime obscured	Conflictual nature of crime recognized
Wounds of offender peripheral	Wounds of offender important
Offense defined in technical, legal terms	Offense understood in full context: moral, social, economic, political

(Source: Zehr, 2015)

In restorative practice, accountability is very important factor accountability, in the restorative practice method, the action taken by the person that causes the harm does not generate only punishment it also generates a responsibility, a responsibility of repairing the harm within the person affected by their actions. When wrongdoing occurs, individuals must be held responsible in ways that reflect the logical consequences of their conduct. This accountability is comprehending and admitting wrongdoing and taking actions to rectify the situation. Individuals must hold society responsible by assisting in identifying and meeting their needs. Similarly, the greater community must heed the needs of individuals that cause harm, aiming not just to restore but also to transform. Accountability is multi-layered and transformative. (Zehr, 2015).
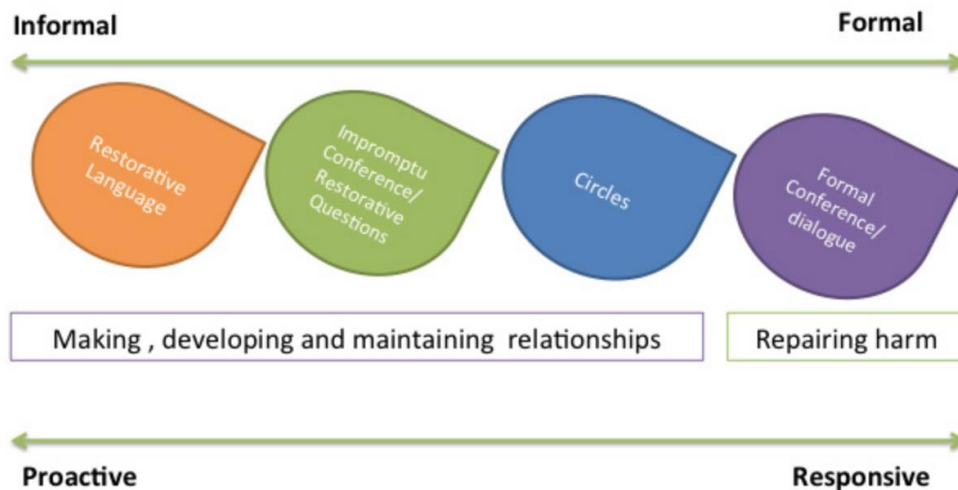
Retributive Lens	Restorative Lens
Wrongs create guilt	Wrongs create liabilities and obligations
Guilt is absolute, either/or	Degrees of responsibility
Guilt is indelible	Guilt is removable through repentance and reparation
Debt is abstract	Debt is concrete
Debt is paid by taking punishment	Debt is paid by making right
Debt is owed to society in the abstract	Debt is owed to victim first
Accountability as taking one's "medicine"	Accountability as taking responsibility
Assumes behavior chosen freely	Recognizes difference between potential and actual realization of human freedom
Free will or social determinism	Recognizes role of social context as choices without denying personal responsibility

(Source: Zehr, 2015)

As we can realise restorative practice is a more transformative approach the focus is not only on the punishment itself or just proving who is guilty or not. It is about responsibility for actions that affect directly other people. The differential of this type of approach is making people that cause harm think about their actions and giving them a chance to transform this into a positive way and also contribute to avoiding further future harmful actions.

1.3.2. Restorative Practice effectiveness on physical grounds.

As shown before restorative practice is an approach to addressing issues and like all approaches, it has a certain technique how to implement the approach effectively and successfully. Project, n.d. states that is a framework for developing, sustaining, and deepening relationships, as well as reacting to conflict via real discussions, that promotes a common understanding of what happened and how to make things right. This framework is composed could be mainly composed of restorative languages, restorative questions, restorative circles and dialogue.

Informal   Formal

Restorative Language   Impromptu Conference/ Restorative Questions   Circles   Formal Conference/ dialogue

Making , developing and maintaining relationships   Repairing harm

Proactive   Responsive

Adapted from ©IIRP EUROPE 2013

(Source: Limerick Restorative Practice Project, n.d.)

The language we use to communicate with one another is a critical component of the development and integration of restorative practices. This applies to both ways we communicate with one other and also how we listen to each other in the setting of restorative methods. While the restorative questions are reflective inquiries that encourage people to consider how their behaviour has affected them and others. They may be utilised in any situation and tailored to every child's or young people's emotional literacy as well as comprehension.

In relation to restorative circles, they foster an environment and culture in which all participants are treated equally, share accountability for results, and establish a sense of connectivity, respect, and comfort. And at the same time, the restorative Conference is a systematic gathering of individuals who have inflicted harm, those who have been affected, and those who might be impacted or can assist those participating. (Project, n.d.)

The effectiveness of restorative justice has been proven over the years for a closer analysis of The Hate Speech Restorative Conference between a traveller family and a Garda published on the website Restorative Justice: Strategies for Change, n.d. Will be used as an example. This example is a travelling family (Michel, Mary, and children) who's been on the housing list for ten years and had finally been allocated a house in a housing state in the nearby town.

So, the family was getting ready to move, even though the inhabitants of the housing state had not accepted it so well, and when people found out that the family was about to move, people decided to start a petition to try to stop them, and then they began to try to include as many people as possible to sign the petition. Many individuals became interested that one of the locals formed a WhatsApp group that includes people who did not previously live in the state. It happened when one of the family's kids (Michel Jr.) was inadvertently introduced to the group. When this occurred, Michel Jnr was shocked by the substance of the derogatory remarks directed against his family and the Travellers group. One person, a Garda, made a very destructive and threatening statement that had a particularly negative impact on the family.

Michael and his family were not only dissatisfied with the circumstances, but they also felt threatened and didn't want to move. Even members of the family who had no problems with the law or criminal records were concerned that their safety might be jeopardised. The family attempted to find another residence since they did not feel secure returning to that house. However, the council stated that it was not allowed because of housing scarcity. A conference was organized, and Michael, Mary, and the Garda were all present there. Michael and Mary were curious about why the Garda would not want them to relocate to the state. The Garda stated that he was subjected to a great deal of stress from his wife's parents, who live in the state. The garda apologised and stated that he acted impulsively and was unaware of how much damage he had caused. He apologised several times to the family, saying he didn't want to create so many problems.

Michael also asked if the garda had something against the travelling population in general, to which the garda replied that he didn't know much about them just that there is a sense of dread among the people who live in the state due to a stereotype which the media had established. Michael then proposed to the garda that they undertake some volunteer work with the traveller community to get to know them better, and the garda accepted. Finally, the conference was a success, and the garda assisted the family in finding another property in a different region. In addition, the garda has been helping with the local traveller group since the conference, and he stated that "he never understood much about traveller culture or its customs, but now he embraces them."

This specific example can illustrate how the use of restorative practice addressing hate speech can be beneficial for not just the parties involved but also for the whole community.

When compared to more typical criminal justice programmes, a meta-analysis indicated that restorative justice was more successful in enhancing victim and offender satisfaction, lowering offender recurrence, and boosting offender compliance with reparations. (Latimer et al., 2005)

The examples are not limited to that one, this specific case was used in this study because it is a case that deals with the same subject of this study. However the website Restorative Justice: Strategies for Change, n.d. bring plenty of examples of how restorative practice have been successful in Ireland. According to them:

Across Ireland, State and non-State actors operating in the criminal justice arena have adopted restorative principles and practices in their work. This includes the day-to-day use of restorative language and relational approaches to practice, as well as their informal and formal use of restorative circle processes to help build relationships, consult colleagues and citizens, structure challenging conversations and respond to conflict. These case studies illustrate just some of the innovative and progressive work taking place in Ireland.

Furthermore, an analysis made by the United Kingdom government showed that 85 per cent of victims who took part in restorative justice programs were satisfied with the process and has determined that restorative justice lowers reoffending by 14 per cent. (Council, n.d.) Not only in the United Kingdom but restorative justice practices are used and known worldwide, and some countries use this practice even to address the harm caused by wars.

2. RESEARCH METHODOLOGY

Introduction

It is vital to investigate the effects of hate speech distributed on social media today, as well as the reasons and current mechanisms in place to address the issue. The history and growth of hate speech on social media will be critically examined in this work for the reader to get a great insight into the problem, and the efficiency of the restorative practice applied to address hate speech on social media will be evaluated by collecting reliable research using a questionnaire answered by social media users to generate an accrued study.

2.1. Philosophies

According to the onion model presented by Saunder et al. (2019) Research philosophy is a set of ideas and assumptions regarding the evolution of knowledge. This may seem complex

however, it is exactly what is done when embarking on research that means gaining knowledge in a certain topic. Your knowledge development may not be too drastic, but even tackling a specific problem in a specific organisation requires the production of new information. In this context ontology, epistemology, and axiology are the three essential components.
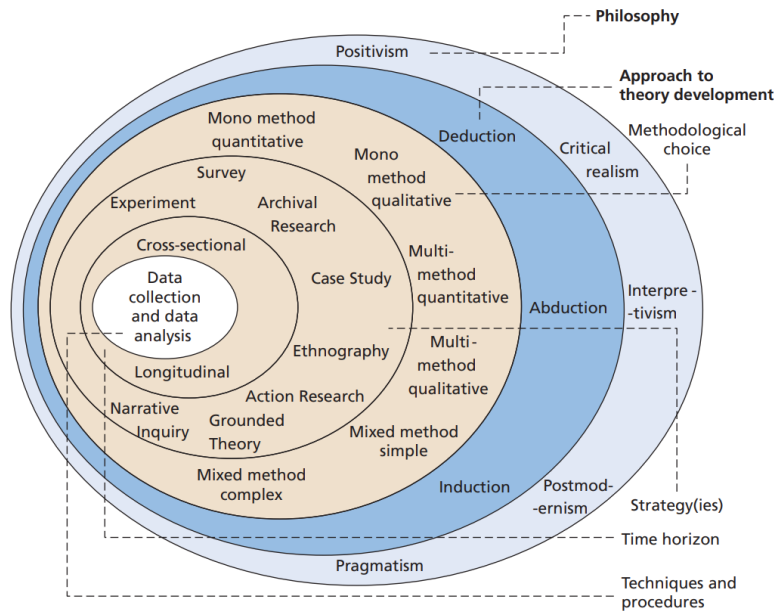
Figure 4.1 The 'research onion'
Source: ©2018 Mark Saunders, Philip Lewis and Adrian Thornhill

(Source: Saunders et al., 2019)

This research adopted a relativist ontology because we will be looking at social media users' viewpoints, beliefs, attitudes and behaviour patterns, and we anticipate a range of realities among them. As a result, because the research will be done with social media users, we'll employ an EMIC epistemology.

Ontology is the assumptions we make about what it means for something to exist that are central to ontology, which is concerned with reasoning about the nature of reality. Maybe the most crucial topic dealing with ontology would be that the social reality we examine should be viewed as existing objectively, externally to observers, or as being made real by human activity and the meaning that observers attribute to them. Our first goal as social researchers should be to comprehend reality. Various ontological perspectives define reality in multiple ways. Therefore, the ontological assumptions we take affect what we aim to comprehend through study. Researchers may construct research studies that are most

successful at capturing the reality we aim to comprehend by knowing ontology and hence our ontological premises. (Emma Bell et al., 2022)

Subjectivism integrates ideas from the arts and humanities, arguing that social reality is formed by social people's views and subsequent actors. Subjectivism accepts nominalism (also known as conventionalism) ontologically. (Saunders et al., 2019)

Table 4.1 Philosophical assumptions as a multidimensional set of continua

Assumption type	Questions	Continua with two sets of extremes		
		Objectivism	⇔	Subjectivism
Ontology	• What is the nature of reality?	Real	⇔	Nominal/decided by convention
	• What is the world like?	External	⇔	Socially constructed
	• For example:	One true reality (universalism)	⇔	Multiple realities (relativism)
	– What are organisa-tions like?	Granular (things)	⇔	Flowing (processes)
	– What is it like being in organisations?	Order	⇔	Chaos
	– What is it like being a manager or being managed?			
Epistemology	• How can we know what we know?	Adopt assumptions of the natural scientist	⇔	Adopt the assumptions of the arts and humanities
	• What is considered acceptable knowledge?	Facts	⇔	Opinions
		Numbers	⇔	Written, spoken and visual accounts
	• What constitutes good-quality data?	Observable phenomena	⇔	Attributed meanings
	• What kinds of contribu-tion to knowledge can be made?	Law-like generalisations	⇔	Individuals and con-texts, specifics
Axiology	• What is the role of val-ues in research? Should we try to be morally-neutral when we do research, or should we let our values shape research?	Value-free	⇔	Value-bound
	• How should we deal with the values of research participants?	Detachment	⇔	Integral and reflexive

(Source: Saunders et al., 2019)

Contrastingly according to Phair & Warren (2021) epistemology is concerned with "how" we get knowledge and learn to comprehend things, in other terms, how we determine what reality is and the limitations of our understanding. We should examine the concept of epistemology and how it informs study design by examining the relationship between a subject and an object. Objectivist epistemology holds that reality exists outside of, or apart from, the individual mind. Objectivist research is beneficial in terms of dependability and external validity. The assumption that objective 'truth' exists and is waiting to be revealed is rejected by constructionist epistemology. Instead, 'truth,' or meaning, emerges through our involvement with the reality of our environment. That is, there is no real-world independent of human action or symbolic language. Constructivism research adds value by developing contextual understandings of a certain topic or situation. (Moon, Blackman, 2017)

In relation, emic and etic are anthropological terms that define two main methods of examining language and society. The emic–etic duality has affected how areas as diverse as psychological science, consumption patterns, organisational science, and intercultural communication investigate cultural systems. In the context of ethnographic fieldwork, the

phrases may refer to diverse research strategies. According to American University (n.d.) "The etic perspective is the outsider's perspective, the perspective that we have of a project's parameters." and "The emic perspective is the insider's perspective, the perspective that comes from within the culture where the project is situated." In this case, the research follows the Emic method because will be analysed from the inside perspective of the social media users.

2.2. Approaches

To assess the study to be conducted with social media users from multiple platforms, an inductive technique was used. Which includes developing hypotheses via research rather than beginning a project with a thesis as a basis. (Phair, Warren 2021)

Unlike the deductive technique, the inductive approach allows you to develop your theory rather than using one that already exists. This clearly distinguishes the two ways. The inductive method is distinguished by a shift from the specific to the universal. Furthermore, this approach is frequently employed in qualitative research. This strategy, however, may also be employed successfully within positivist approaches, in which data is processed first and noteworthy patterns are used to influence the development of outcomes. (Thesismind, 2019)

2.3. Strategies

According to Saunders et al. 2019, A research strategy specifies how the researcher will carry out the study based on the objective of the project. Thesismind (2019) A variety of methodologies, such as experimental research, case study, descriptive study, interviews, surveys, or a systematic literature review, can be used in the plan.

For this study, the strategy chosen is the Survey in form of a questionnaire containing 22 close questions to try to extract qualitative and quantitative data on the experiences of social media users about hate speech. The essence of the survey technique is "questioning people about a topic or subjects and then reporting their answers." In business studies, the survey technique of primary data collecting is used to test theories, reflect social attitudes, determine the degree of customer satisfaction, undertake segmentation research, and for a

variety of other objectives. The survey approach could be used in either quantitative or qualitative research. (Thesismind, 2019)

Survey research is a quantitative method that uses self-report measurements of a limited group of people. It is a versatile method that may be used to investigate a wide range of basic and practical research problems. This category includes articles that discuss utilising a survey approach. (Irati Vizcarguenaga-Aguirre, 2020)

2.4. Choices

This topic determines how many different data categories (qualitative or quantitative) you will employ in your investigation. There are three methods available: mono, mixed, and multi-method. (David Phair, 2021)

When utilising a mono method strategy, you must collect only one sort of information, which might be quantitative or qualitative. The two cannot be combined. While a mixed method enables researchers to integrate quantitative and qualitative methodologies in a study to provide a precise collection of data. In addition, the multi-method approach is comparable to the mixed method in that both incorporate quantitative and qualitative methods into research. Although they look the same, they are not identical. Multi-method does not integrate methodologies to establish a specific collection of data, whereas mixed-method does. In this research, it was used the multi-method as some questions will be analysed statically and others will be analysed critically to understand patterns, behaviours and opinions. (Thesismind, 2019)

2.5. Time Horizon

According to Bryman (2012), The time horizon specifies the amount of time necessary to complete the project task. The research onion specifies two sorts of temporal horizons that can be used cross-sectionally and longitudinally.

The cross-sectional time frame has already been determined, and data must be collected inside it. When the inquiry is focused on the study of certain phenomena at a given moment, this is employed. Whereas longitudinal information gathering refers to collecting data

periodically over a lengthy period and is employed when evaluating change through time is a significant component of the research. (Goddard & Melville, 2004) (Thesismind, 2019)

In the present research, because a longitudinal study may take several years, which would be too long for this research, a cross-sectional time horizon was chosen. Additionally, due to the nature of the subject of the research being very variable and quickly changeable over time the use of the cross-sectional approach proves to be more suitable.

2.6. Target population

Reference (2020) states that members of a group that a researcher is interested in researching constitute the target population. The study's findings apply to this demographic because all of them share important characteristics. A target population is just the group of people you've chosen to examine or study. A subset of the target population is referred to as a sample population. Because surveying or analysing the complete group is nearly impossible, sample populations are frequently utilised in research. Typically, sample populations are a group of people that closely resemble the demographics of your selected target audience. Different forms of sampling have different data collection methodologies. (Windham, n.d.)

The population consists entirely of social media users, but because this is such a large and varied group, it would never be possible to quantify every member inside the group, making the possibility of sampling impossible. As a result, the users' survey, will be employed non-probability sampling and a predetermined sampling of a minimum of 100 social media users.

2.7. Data collection and data analysis

A questionnaire via an online survey was used to collect data for this study using google forms. The questionnaire was composed of 22 questions, being 2 of them demographic questions (1. What's your gender, 2. How old are you?) and 20 questions based on the target population's experiences, behaviours and knowledge and 15 of them use the Likert scale form. Likert Scale questions are a type of closed question that is extensively employed in popular opinion research. Psychometric testing is used to assess beliefs, attitudes, and opinions. Statements are used in the questions, and respondents indicate how much they

agree or disagree with each assertion. With Likert Scale questions, a scale of 0-10 is often supplied, while shorter scales may be conceivable. (SmartSurvey, n.d.)

The survey is entitled "Restorative Practice as an alternative to address hate speech on social media." And the objective of this survey is to examine the effects of hate speech distributed on social media, as well as to determine what techniques platforms are employing to address the issue and if such efforts are effective. Additionally, the research will be useful in determining whether a different approach, such as restorative practice, may be beneficial in addressing the issue.

For data analysis, the questionnaire is classified as quantitative and qualitative. Qualitative research is concerned with definitions and meanings, whereas quantitative research is concerned with statistics and numbers. It is possible also to develop a mixed methods research study that incorporates components of both. (Thesismind, 2019) In this case, the research looks for measuring and interpreting, once some questions have the aim to analyse statically the incidence of the hate speech phenomena other questions aim to interpret the user's experience, opinions and knowledge about the subject.

2.8. Ethics

The beginning of the survey has brief information about what is the subject of the research and all participants that have taken the research was informed that the survey was for academic purpose only and was also informed that aiming to protect confidentiality the survey did not have any personal information questions able to identify who was answering. In the explanation, it was stated that by submitting their answers the participants were agreeing to the survey and allowing their answers to be analysed and used as part of the research. In addition to that, the questionnaire was also optional, and any participant can withdraw at any moment.

2.9. Research Limitations

One of the research's limitations is that the topic is very changeable, therefore it will take some time to develop and execute ideas and by the time that the ideas are developing things can change quickly and might shift perspectives. Another concern is that because the subject being described is relatively new, it may have been difficult to acquire theoretical

references. And, because this is a complicated and deep-rooted topic, the research may not yield a clear solution. Furthermore, because the poll was conducted with a small number of individuals, this option may not always represent the ideas, attitudes, or behaviours of the whole society.

## 3. PRESENTATION OF DATA

Introduction

As explained previously the survey consisted of 22 questions, which 2 of them were demographic questions based on their gender and age and 20 closed questions, 15 of them were designed using the Likert scale model and all the 20 questions were related to the participant's knowledge, experience and behaviours based on what they had experienced online and what their level of knowledge about restorative practice approach.

3.1. DEMOGRAPHIC QUESTIONS

According to SurveyMonkey (n.d.) Demographic questionnaires are used to obtain basic information on respondents to help survey designers determine where each person fits in the larger population. Demographic data helps you to better grasp an audience's background characteristics, such as age, race, ethnicity, income, job position, marital status, and so on. In this research 2, demographic questions were asked which were their gender and their age. In the survey 85.6% were female and 14.4% as we can analyse in figure 1:
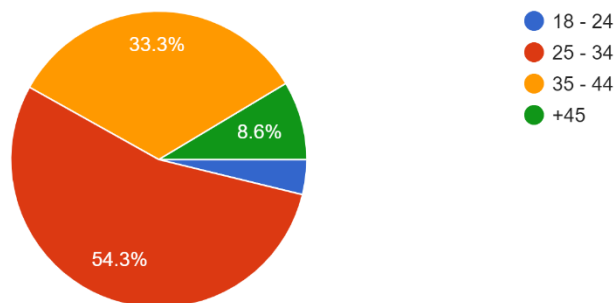
Figure 1 – What is your gender?

1. What is your gender?
104 responses

● Male
● Female
● Prefer not to say

85.6%

14.4%
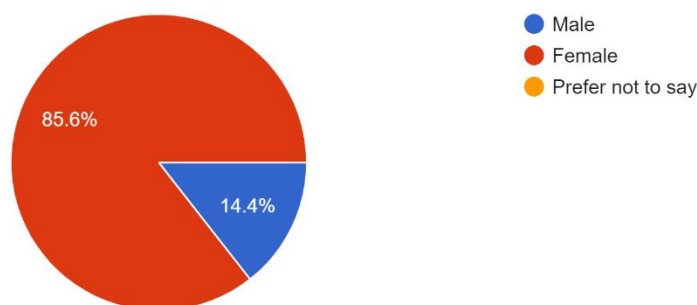
In the second question what was asked was the age of the participants and it was provided 4 range of age options, which 3.8% were 18-24, 54.3% were 25-34, 33.3% were 35-44 and 8.6% were over 45 years old.

Figure 2 – How old are you?

2. How old are you?
105 responses

- 18 - 24
- 25 - 34
- 35 - 44
- +45

3.2. CLOSED QUESTIONS

This research was composed of 20 closed questions and 3 questions has only yes or no as an option (Have you ever been a target of hate speech on social media?, Did you ever use any tools provided by social media to report hate speech?, Have you ever heard about restorative practice?), 2 has yes or no and maybe, which the thirteen was a condition to answer the number fourteen and the number nineteen a condition to answer the number twenty (If yes. Do you consider the restorative practice could be an effective approach against hate speech on social media?, Would you consider the restorative practice a valuable way to address the issue?) and the other 15 were based in Likert scale form where were used with the aim of to assess individual's views, attitudes, and opinions.

In question thirteen, it was asked if the participant had been a target of hate speech on social media and had the purpose of statically measuring people that had experienced hate speech. In this question, 84.8% answered No and 15.2% answered yes.

Figure 3 – Have you ever been a target of hate speech on social media?

13. Have you ever been a target of hate speech on social media?
105 responses
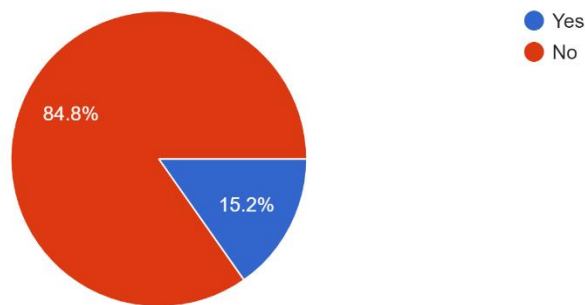
In question sixteen it was asked if the participants have ever used any tools available on social media to report hate speech. In this question, the objective was to analyse if people are likely to engage in hate speech that they perceive. 67.3% of the participants answered Yes and 32.7% said No.

Figure 4 – Did you ever use any tools provided by social media to report hate speech?

16. Did you ever use any tools provided by social media to report hate speech?
104 responses
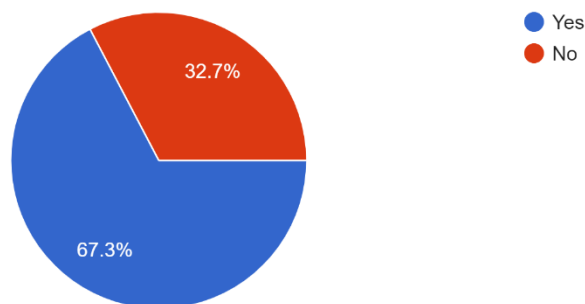
Question nineteen was to simply measure how popular restorative practice on social media is among users. And it was a conditional question that people that answered yes were conducted to answer question twenty. To this question, 59% said No and 41% said Yes.

Figure 5 – Have you ever heard about restorative practice?

19. Have you ever heard about restorative practice?
105 responses

- Yes
- No
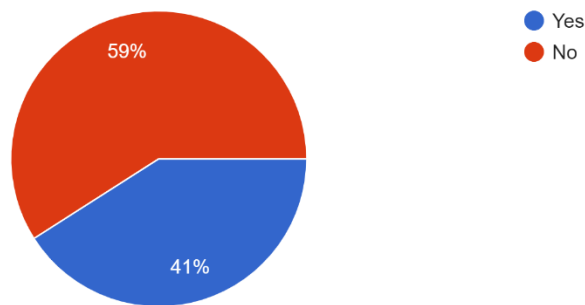
As explained above question twenty was exclusively for people that had answered Yes to question nineteen, once question twenty is to analyse people's opinions even if they were minimal knowledgeable about restorative practice's effectiveness against hate speech. Of which 69.8% said Yes, 27.9% Maybe and 2.3% said No.

Figure 6 – If yes. Do you consider the restorative practice could be an effective approach against hate speech on social media?

20. If yes. Do you consider the restorative practice could be an effective approach against hate speech on social media?
43 responses

- Yes
- No
- Maybe
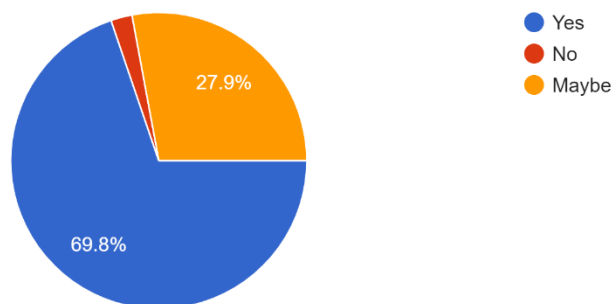
The last question has the aim to evaluate If the participants would consider the restorative practice approach to address hate speech in the social media environment. 57.8% said Maybe and 42.2% answered Yes and no 2.3% answered No.

Figure 7 – Would you consider the restorative practice a valuable way to address the issue?

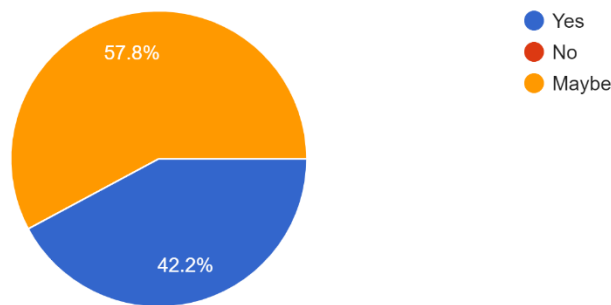22. Would you consider the restorative practice a valuable way to address the issue?
102 responses

3.3. LIKERT SCALE QUESTIONS

As stated before, in this questionnaire 15 questions were designed attending the Likert scale model which is very popular in research where researchers want to measure opinions, knowledge and/or experience more subjectively and analytically. A Likert rating scale is a measurement tool that is used in surveys to generate questions. It is among the most common forms of survey questions. Participants in a Likert scale survey do not pick simply between 'yes/no,' but rather between 'having to agree' or 'simply disagree' on a single survey topic. (QuestionPro, n.d.)

In the third question, it was inquired how frequently participants use social media. 62.9% of them answered Daily, 33.3% replied Hourly, 2.9% Weekly and 1% Monthly.

Figure 8 – Do you use social media often?

3. Do you use social media often?
105 responses
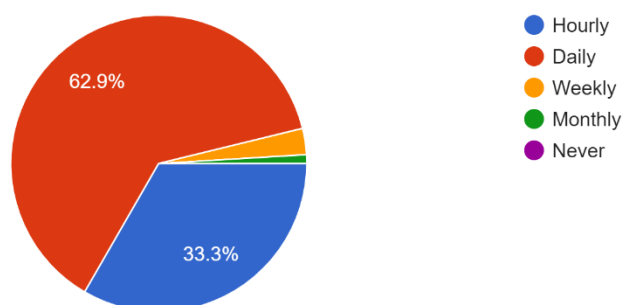
Question four asked the participants which social media they use the most. It was given four options in the question (Facebook, Instagram, TikTok and Snapchat) however was also given to the participants the space "Other" be free to add any other social media that was not included in the options. For this question, 81.9% answered Instagram, 12.4% Facebook, 2% TikTok, 1% said WhatsApp and Instagram, 1% Twitter and other 1% replied Twitter what be realised as probably a typo and another 1% said LinkedIn.

Figure 9 – Which of the following social media do you use the most?

4. Which of the following social media do you use the most?
105 responses

- Facebook
- Instagram
- TikTok
- Snapchat
- Twitter
- WhatsApp and Instagram
- Twiter
- Likedin
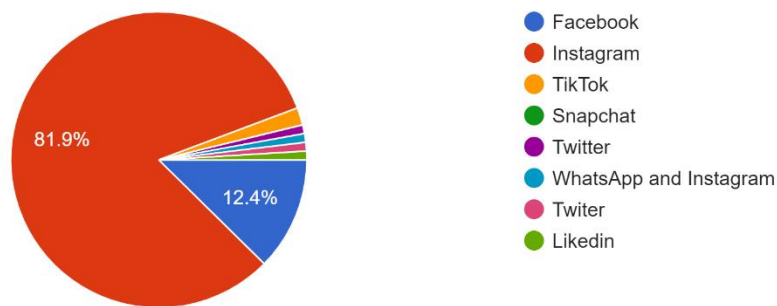
In question, five participants were asked about the frequency that they witness hate speech in a specific area on the social media that they use the most and for this question, it was about sexist speech. 44.8% said Sometimes, 31.4% Often, 12.4% Always, 10.5% Rarely and 1% said Never.

Figure 10 – How often do you witness sexist speech on social media that you use the most?

5. How often do you witness sexists speech on social media that you use the most?
105 responses

- Always
- Often
- Sometimes
- Rarely
- Never
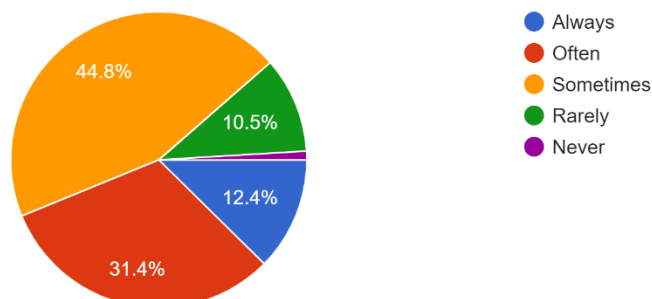
Following the same line of question five but on a different subject, question six inquired the participants how often they witness racist speech on social media that they utilise the most. 49,5% replied Sometimes, 23,8% said they witness this type of speech Often, 16,2% Rarely, 8,6% Always and 1,9% said Never.

Figure 11 – How often do you witness racist speeches on social media that you use the most?

6. How often do you witness racist speeches on social media that you use the most?
105 responses
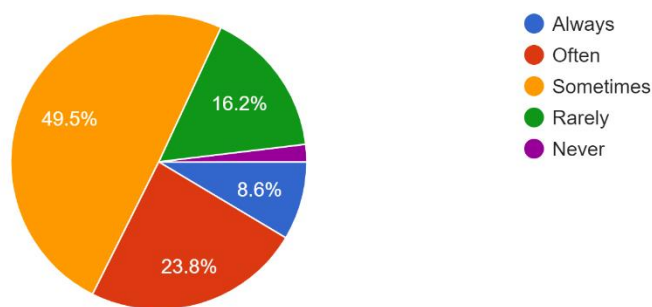
For question, seven social media users were asked how frequent they witness hate speech related to religious intolerance on social media. 31,1% said Sometimes, 27,6% Often, 23,8% Rarely, 7,6% Always and 3,8% Never.

Figure 12 – How often do you witness religious intolerance speeches on social media that you use the most?

7. How often do you witness religious intolerance speeches on social media that you use the most?
105 responses
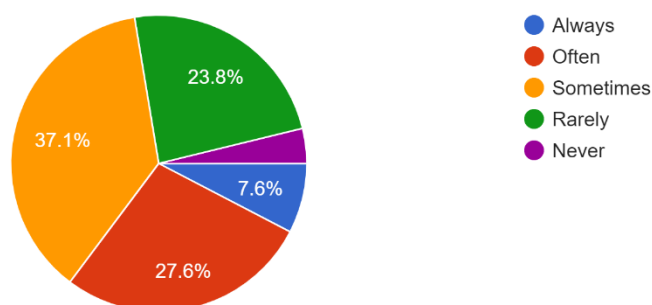
In question eight in the same model as the previous three questions participants were asked how often political intolerance witnessed on the social media was that they use frequently. For this question 41,9% of the participants answered Often, 38,1% replied Always, 14,3% Sometimes, 3,8% Rarely and 1,9% stated Never.

Figure 13 – How often do you witness political intolerance speeches on social media that you use the most?

8. How often do you witness political intolerance speeches on social media that you use the most?
105 responses

Finally, question nine was the last question that follows that pattern of the previous ones, aiming to analyse how frequent participants view hate speech about homophobia and the LGBTQIA+ community. In this question 33,3% answered Often, another 33,3% answered Sometimes, 16,2% said Rarely, 13,3% Always and 3,8% Never.

Figure 14 – How often do you witness homophobic or LGBTQIA+ intolerance speeches on social media that you use the most?
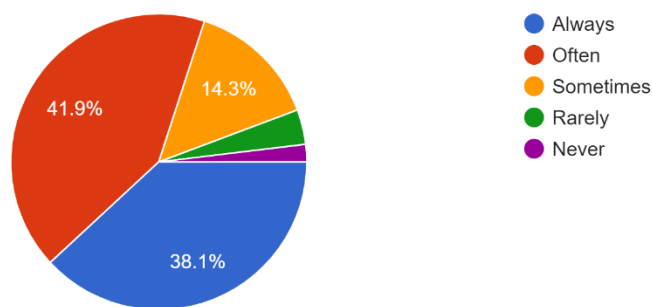
9. How often do you witness homophobic or LGBTQIA+ intolerance speeches on social media that you use the most?
105 responses
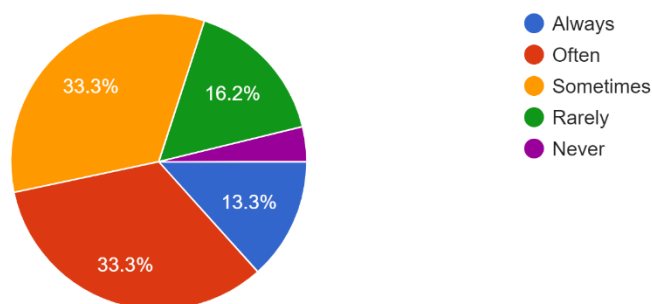
On question ten to analyse the participant's behaviour when they evidence hate speech, they were asked how likely they were to engage in social media against the speeches described in the previous five questions. 25,7% responded Possibly, 21% Probably, 21,9% Probably Not, 16,2% Not and 15,2% responded that they definitely would engage on social media against the hate speeches described before.

Figure 15 – How likely are you to engage in social media against the speeches described above?

10. How likely are you to engage in social media against the speeches described above?
105 responses

- Definitely
- Probably
- Possibly
- Probably not
- Definitely not
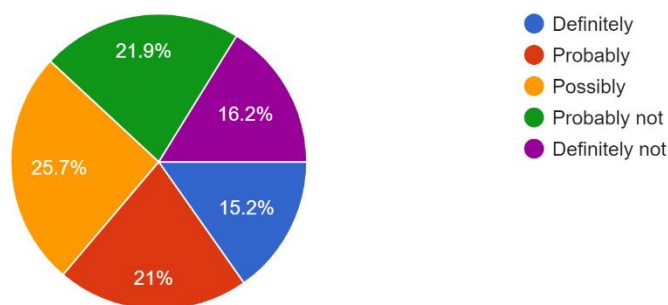
Question eleven was designed to assess how people judge the importance of discussing speech. 62,5% judge it to be very important to discuss hate speech, 16,3% think is only important, 14,4% judge it to be not important and 6,7% believe that is Somewhat important.

Figure 16 – How important is discussing hate speech?

11. How important is discussing hate speech?
104 responses

- Very important
- Important
- Somewhat important
- Not important
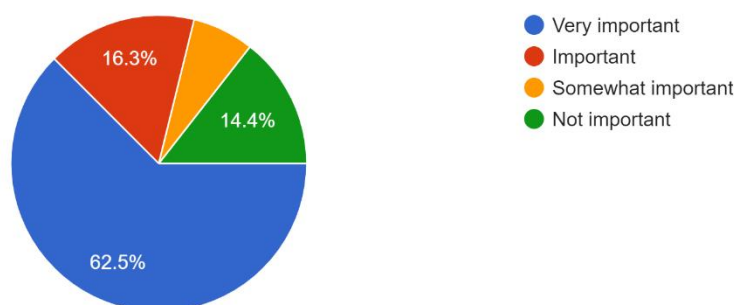
In the same direction as question eleven, question twelve has the objective to assess how important people judge free speech. In this question 61,9% answered that is Very important, 30,5% for just Important and 7,6% answered that is Somewhat important.

Figure 17 – How important is free speech?

12. How important is free speech?
105 responses

- Very important
- Important
- Somewhat important
- Not important
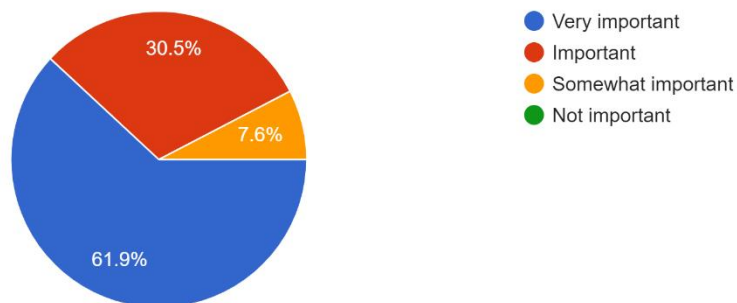
Question fourteen was a pre-determined by question thirteen which means that people that had answered Yes to the question "have you ever been a target of hate speech on social media" were led to question fourteen for the people that said No the questionnaire was designed to skip question fourteen automatically for those participants. Once the aim of question fourteen is to evaluate how hate speech on social media experienced by participants affected them. Of 15.2% of the participants that answered Yes to question thirteen 31.3% of them responded that this fact had affected them Very little, while 25% responded Quite well, 18.8% Fairly well, 12.5% Very well and 12.5% not at all.

Figure 18 – If yes. Did this event affect you?

14. If yes. Did this event affect you?

16 responses

Legend:
- Not at all
- Very little
- Fairly well
- Quite well
- Very well

Pie chart values: 25%, 12.5%, 18.8%, 12.5%, 31.3%
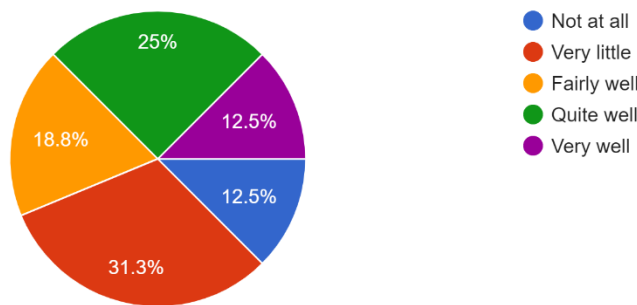
Question fifteen was also linked to question fourteen and it was also designed to know how the experience about the support of the social media was, consequently just replied to this question who said Yes to question thirteen. In this question 58.8% said None, 23.5% said they had considerable support from the social media, 11.8% answered Slightly and 5.9% had Great support from social media.

Figure 19 – Did you have any support from the social media community?

15. Did you have any support from the social media community?

17 responses

Legend:
- None
- Slightly
- Considerable
- Great

Pie chart values: 11.8%, 23.5%, 58.8%
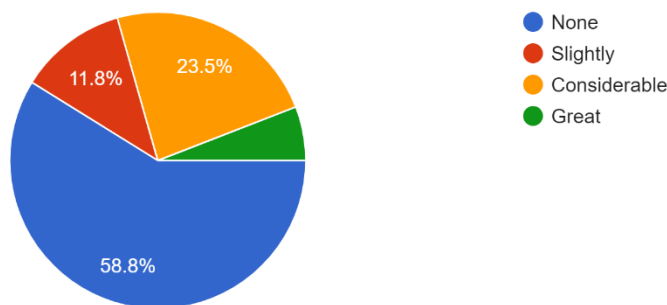
Question seventeen were planned to analyse participants' opinion about the importance of social media in providing tools and internal measures to report and tackle hate speech. To this question, 73.3% answered that is Extremely Important, 21.9% consider it Very Important, 4.8% believe it to be Moderately important, and no responses were marked for the options Slightly Important and Not important at all.

Figure 20 – What is the importance of having tools and internal measures inside social media to cope with hate speech?

17. What is the importance of having tools and internal measures inside social media to cope with hate speech?

105 responses

- ● Extremely Important
- ● Very Important
- ● Moderately
- ● Slightly important
- ● Not important at all
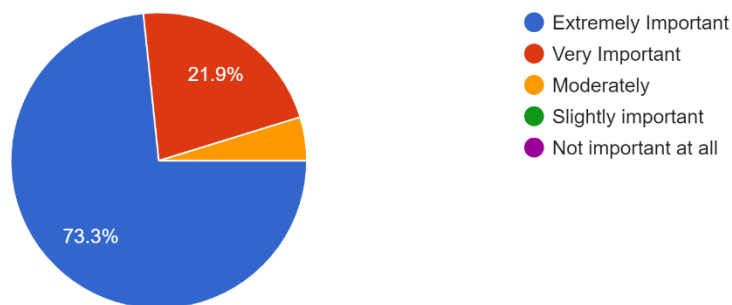
Question eighteen was asked the participant's opinion if they think that social media content review and internal policies ended up reflecting negatively on free speech. For this question 32.4% of the participants were Neutral, 22.9% Agreed that content review and internal policies do interfere with free speech, 21.9% of the participants Disagree that content review and internal policies would be prejudicial to free speech and 4.8% Strongly Disagree that those factors interfere in free speech.

Figure 21 – In your opinion social media content review and internal policies interfere with free speech?

18. In your opinion social media content review and internal policies interfere with free speech?

105 responses

- ● Strongly agree
- ● Agree
- ● Neutral
- ● Disagree
- ● Strongly Disagree
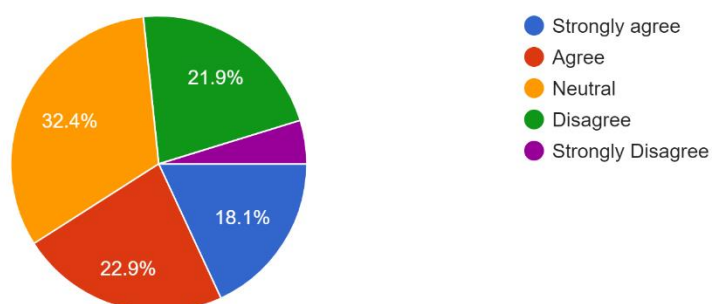
Finally, question twenty-one were to evaluate the level of knowledge of the participants taking the questionnaire about the restorative practice approach. For this question 55.2% of the participants answered that they were Not knowledgeable about it, 20% were Somewhat

knowledgeable about it, 19% were Knowledgeable about and 5.7% were Very knowledgeable about it.

Figure 22 – What is your knowledge about restorative practice?

21. What is your knowledge about restorative practice?
105 responses
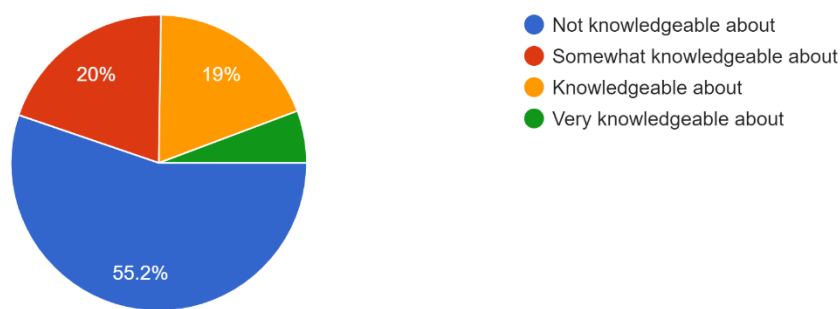
4. DATA ANALYSIS AND FINDINGS

Introduction

This chapter will analyse the data collected through the questionnaire to provide the findings for the research objectives. In total one hundred and five people responded to the questionnaire and these findings include the gender and age distribution, the social media distribution, the hate speech data, the free speech data and the restorative practice data. Which will be critically analysed to evaluate if the restorative practice approach to address hate speech on social media would be viable.

4.1. Gender and age distribution

According to the demographic question was possible to find out that the majority of the participants were women 85.6%, while men were represented by 14.4%, indicating that women were more active on social media and more willing to engage with the subject. Regarding the age, most of the people that took the questionnaire were between 25 to 44 years old (54.3% were 25-34 and 33.3% were 35-44) so it is possible to say that women between their early twenties and early forties might be more likely to have interest and engage in this topic.

4.1. Social media distribution

In this topic, it is possible to observe that is highly and constantly used, once most the people respond that they use social media daily (62.9%) and another big part use social media hourly (33.3%) and the rest uses at least weekly (2.9%) and only 1% use monthly. According to this result, it is possible to assume that social media is constantly present in the participant's daily life which is possibly also reflected in the amount and type of content that they are exposed to daily. In addition to that engagement in general in social media increase if people spend more time with it. In this topic also were possible to ascertain that the most popular social media is Instagram (81.9%) while only another small portion uses Facebook more often (12.4%) and only a small portion stated that use other social media such as TikTok (1.9%), WhatsApp (1%), Twitter (1%) and LinkedIn (1%). Therefore, the age factor could have influenced this topic once most of the participants were 25-44 and the oldest

social media in the options were Facebook founded in 2004 and Instagram founded in 2010, while TikTok is a newer social media founded only in 2016. (University, n.d.)

In addition to that people were asked if they had ever been a target of hate speech on social media and if they did have any type of support from the social media and the most of them answered that they had no support from the social media (58.8%) and some people had considerable support (23.5%) while some of them had only a slightly support (11.8%) and only a small portion had great support (5.9%). In contrast with these results, most of the people stated that they already had used tools provided by social media to report hate speech (67.3%) and just a few people said they have never used any of these tools (32.7%). As most people use this type of tool provided by social media the research also has shown that most of the participants also consider it extremely important (73.3%) the existence of those tools. Another portion of the participants judged it to be very important (21.9%) and another part thinks it is at least moderately important (4.8%). In conclusion, everybody taking the survey thinks it does have some importance to have some type of tools to help cope with hate speech inside the social media environment, even though at different levels of importance.

4.3. Hate speech data

Initially, will be analysed the importance of subject discussion according to the participant's perspective. In this matter more than half of the people answering the questionnaire consider very important the discussion about hate speech (62.5%) and the rest of the participants were quite divided between considering it only important (16.3%) and not important at all (14.4%).

This research found that most of the participants did not ever feel that they were a target of hate speech (84.8%) and only a few of them reported to have already experienced hate speech towards them (15.2%). About how that did have affected their lives most of them reported that this fact has affected them very little (31.3%) in contrast and not very far from this result another portion of the participants reported that this fact had affected them quite well (25%) while another portion stated that they were fairly well affected (18.8%) and the rest of them were equally divided by very well (12.5%) and not at all (12.5%).

About the incidence of the hate speech separated by subject, this research concluded that political intolerance currently has the highest incidence of hate speech on social media, considering that most people witness this type of intolerance always (38.1%) and often (42.9%) and the other part witness this type of intolerance at least sometimes (14.3%) and only a small proportion have never observed (1.9%) or rarely witnessed (3.8%) political intolerance. This reflects the recent emerging popularity in pollical discussions on social media. Social media has become one of the key channels for modern individuals to communicate various political viewpoints. Nowadays users have a wide range of options for calling attention to various civic issues. (Gil de Zúñiga et al., 2014) (Bimber, 2012).

Followed by the political intolerance the second type of hate speech more witnessed in the social media environment is homophobia and LGBTQIA+ for this topic half of the participants were between witnessed this type of intolerance often (33.3%) and sometimes (33.3%) while another few people were between always (13.3%) and rarely (16.2%). After that, Sexist speech is the most common on social media according to the questionnaire most people witness sexist speech at least sometimes (44.8%) and many of them encounter this type of speech often (31.4%) or always (12.4%) and only a few of the participants witness that rarely (10.5%) and only one person had said that never usually does not witness sexist speech (1%).

Following the Sexist speech, it was racist speeches that has the highest incidence on social media. The biggest number of people in this category however usually witness this type of speech sometimes (49.5%), although many of them still have observed this speech often (23.8%) or always (8.6%) while few of them rarely see this type of speech (16.2%) or never see it (1.9%). And with the very similar results, the last one was the religious intolerance also half of the people said that see this type of hate speech sometimes (37.1%) and the second biggest part witness this often (27.6%) and almost the same amount rarely (23.8%) and the rest was divided by always (7.6%) and never (3.8%).

The most interesting results came from the question in which participants were asked if they would engage on social media against the speeches above described. The people were very divided on this question and most of the responses were that they possibly (25.7%) engage against hate speech. Nevertheless, the rest of them were with just a slightly difference some people responded that probably (21%) while not far from that statistic other people answered the complete opposite with probably not (21.9%) while some people

answered that they would definitely (15.2%) engage against hate speech on social media in the other side but not far from that result some other people stated that they would definitely not (16.2%) engage.

4.4. Free speech data

Likely the hate speech this topic will start with people's opinions about how important they judge to be free speech. Most people consider it to be very important (61.9%), another part considers it to be only important (30.5%) and some people consider it at least being somewhat important (7.6%) and contrasting with the hate speech question in this one nobody had marked the option that considers not important at all. Consequently, it is possible to observe that everybody taking this questionnaire judges that free speech has some level of importance.

Participants were also asked if in their opinion the social media content review and social media internal policies interfere with free speech. To this question, most of the participants were neutral (32.4%), and the other biggest part was almost equally divided into two extremes, a little bit more than half of them agreed that those factors do interfere with free speech (22.9%) and a little bit less than half of them disagreed that those measures have interference with free speech (21.9%). However, some strongly agree that this has an interference (18.1%) while just a few of them strongly disagree that this factor restrains free speech (4.8%).

4.5. Restorative Practice data

Initially to analyse that data it will be evaluated how popular the restorative practice approach is among the participants. In this question, most participants reported that they never have heard about restorative practice (59%) while the rest had at least heard about restorative practice (41%). And from this 41% that have at least heard about the approach, most of them believe that restorative practice could be effective to cope with hate speech (69.8%), some of them believe that maybe it could be effective (27.9%) and a small portion holds the opinion that restorative practice would be not effective at all in dealing with this issue (2.3%).

People were also inquired about their level of knowledge about the restorative practice and to corroborate the results about restorative practice's popularity most of the participants

were not knowledgeable about the subject (55.2%) the second biggest part was only somewhat knowledgeable (20%) followed by some participants that stated be knowledgeable about (19%) and only a few of them were very knowledgeable about (5.7%).

Finally in the last question asked the participants needed to express if they would consider the restorative practice as an approach to address hate speech on social media and although some people stated that they would think that the approach would be not efficient to address the issue there were no responses marked for "no" in this question. Most participants reported that they maybe consider the restorative practice to address hate speech (5.8%) while the rest of them reported that they would consider restorative practice as a valuable approach to address the hate speech issue on social media (42.2%). With these results, it is possible to realise that even though the popularity of the approach is not very high people are still open to considering valuable new alternatives to trying to tackle the issue.

5. DISCUSSIONS

Introduction

This chapter will be discussed the findings of the research. It is a very important part of the research that enables the most important data acquired throughout the study to be highlighted, discussed, and compared. The survey questionnaire was comped of 22 questions based on hate speech, free speech, restorative practice and behaviours on social media and it was answered by 105 participants. This is the chapter where it is possible to compare and contrast the survey results with the ideas presented in the literature review.

5.1. Objective 1 – Identify what is the profile of people using social media and which one is the most popular and the incidence of hate speech.

To identify those questions as stated before was used a survey. Closed-ended questions with multiple choice about gender and age were used to identify the profile of the participants and also to identify what is the most popular social media among them. According to the survey, most of them were women between 22-44 years old and the most popular social media among them were Instagram and Facebook. In general, over the years social media platforms gained popularity and have grown quickly. The statistics are staggering: the world population is 7.7 billion, with at least 3.5 billion of us connected. This indicates that social media platforms are utilised by one-third of the world's population and more than two-thirds of all internet users. (Ospina, 2019).

Number of people using social media platforms, 2004 to 2019

Estimates correspond to monthly active users (MAUs). Facebook, for example, measures MAUs as users that have logged in during the past 30 days. See source for more details.

Facebook
YouTube
Whatsapp
WeChat
Instagram
Tumblr
TikTok
Reddit
Twitter
Pinterest
Snapchat
MySpace

2 billion
1.5 billion
1 billion
500 million
0

2004 2006 2008 2010 2012 2014 2016 2019
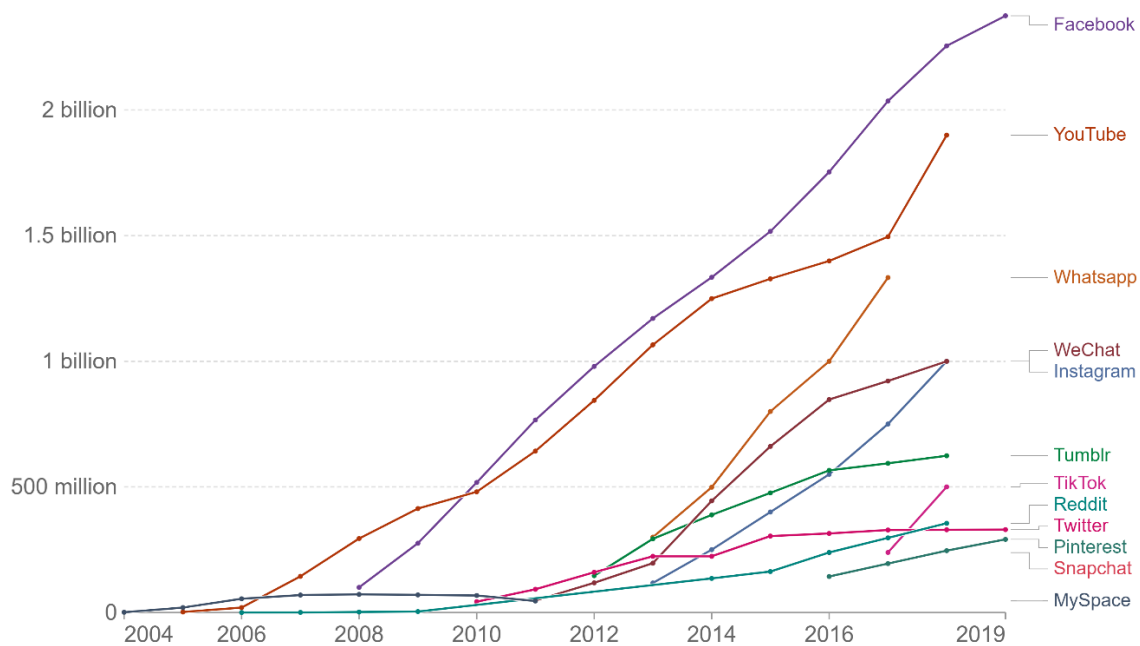
Source: Statista and TNW (2019)

(Source: Our World in Data, 2019)

As explained previously the age factor can highly influence the popularity of participants taking the survey. Is known that in general young people spend more time on social media. However, the age factor can contribute to the popularity on certain social media being higher than on other ones. In this research, the age factor could have contributed to and directly influenced the social media popularity data. To illustrate and use as an example Ospina, 2019 shows that in the graphic. According to her, some social networking platforms are far more popular than others among certain demographics. In the graphic below it is possible to observe that in 2019 Facebook, for example, were more popular with group ages between 25-29 years old and 40-49 years old than among younger people age group that were 18-24 years old.

Use of social media platforms by age group in the US
The share of adults in the United States who say they ever use the following online platforms or social media apps in 2019. This is shown by age group.

18-24 years
25-29 years
30-49 years
50-64 years
65+ years

Youtube — 18-24 years 90%, 25-29 years 93%, 30-49 years 87%, 50-64 years 70%, 65+ years 38%. Avg. across all ages 73%

Facebook — 76%, 84%, 79%, 68%, 46%. 69%

Instagram — 75%, 57%, 47%, 23%, 8%. 37%

Snapchat — 73%, 47%, 25%, 9%, 3%. 24%

Twitter — 44%, 31%, 26%, 17%, 7%. 22%
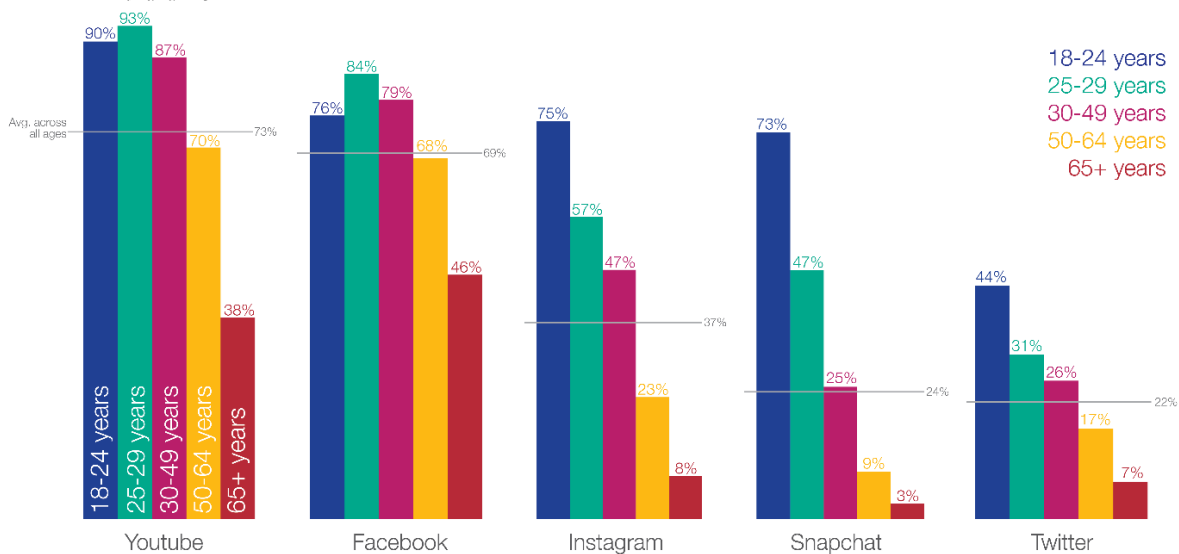
Data source: Pew Research Center (2019).
This is a visualization from OurWorldInData.org, where you find data and research to make progress against the world's largest problems.
Licensed under CC-BY by the author Esteban Ortiz-Ospina.

(Source: Our World in Data, 2019)

However, the graphic can illustrate and create a good example of how demographic questions can be effective for the research. It also needed to bear in mind that dealing and analysing with data that involves the internet, social media and people's behaviours inside these environments it is difficult to gather a very accurate and static data, once in this area, things are very flexible, and the data can change very rapidly in a very small period. Factors such as popularity can shift from one day to another. As explained in the previous chapters this is one of the limitations of this type of research, it might happen that by the time this research finishes the data have changed already.

Another important factor that it needed be pointed out is how much time people expend on social media, once more time people expend online more exposed to a higher content flow, and consequently chances of witnessing or being a target of hate speech increase. According to the survey, most participants are either hourly or daily on social media.

One more time to illustrate that with a similar research Ospina, 2019 reached similar results with research done in the US. According to the researcher

In the US, adults spend more than 6 hours per day on digital media (apps and websites accessed through mobile phones, tablets, computers and other connected devices such as game consoles). As the chart shows, this growth has been driven almost entirely by additional time spent on smartphones and tablets.

52

Daily hours spent with digital media, United States, 2008 to 2018

Average hours per day spent engaging with digital media (e.g. digital images and videos, web pages, social media apps, etc.) The data for 'other connected devices' includes game consoles. Mobile includes smartphones & tablets. All data includes both home & work usage for people 18+.

Our World in Data

Legend:
- Mobile
- Desktop/Laptop
- Other Connected Devices
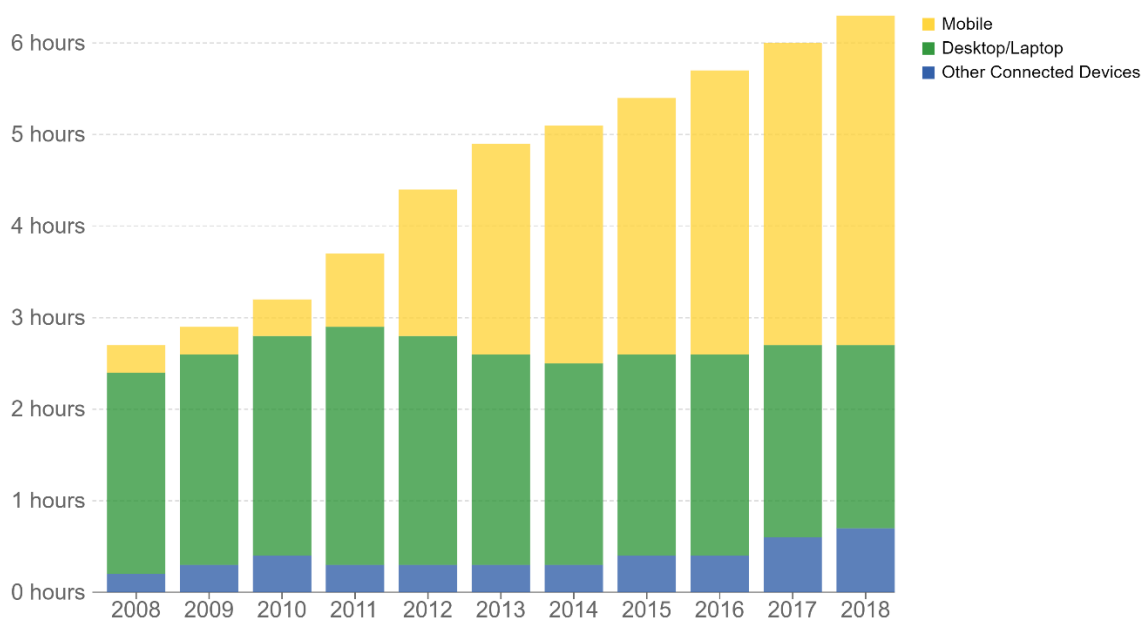
(Source: Our World in Data, 2019)

5.2. Objective 2 – Compare and contrast people's opinions about hate speech vs free speech.

For this topic, as explained before Likert Scale questions were used, due to the belief that this model is the best to evaluate factors such as opinions allowing this type of data to be analysed with more clarity. According to Vinney, 2019 some strong points to use this type of question in research are:

•   A Likert scale enables respondents to choose from a linear set of responses that increase or decrease in intensity or strength. It is a close-ended, forced-choice scale.

•   Widely used in psychological and other social science research today, Likert scales enable researchers to collect data that provides nuance and insight into participants' opinions. This data is quantitative and can easily be analyzed statistically.

•   Likert items often offer response categories on a 1-to-5 scale, but a range of options is possible, including 1-to-7 and 0-to-4 scales or even-numbered scales that typically range from 1-to-4 or 1-to-6.
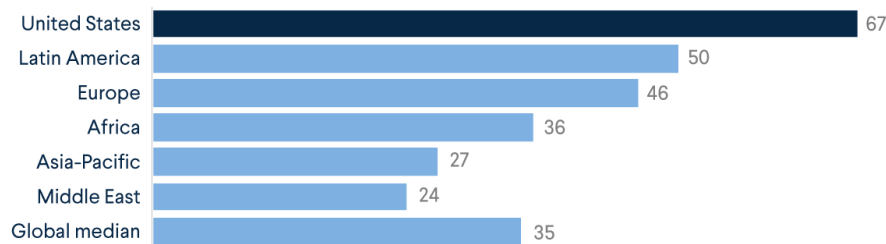
In the present research, people were asked their opinion about the importance of hate speech and free speech both questions were used on a scale of 1-4 to measure the importance of each subject for the participants. The scale started with Very important going until not important. Most of the participants consider both subjects very important and a small part had considered that hate speech is not important.

In the same context were asked to help to corroborate those questions was asked the participants how often they usually witnessed certain types of hate speech online. As explained before in the literature review the types of hate speech exist are infinite so for the purpose of this research, it was focused on five main types of them that include: sexism, racism, religious intolerance, political intolerance, and homophobia which includes LGBTQIA+. According to the results of the survey, the most popular type of hate speech witnessed by the participants is political intolerance, followed by homophobia (LGBTQIA+), sexist speeches, racist speeches and religious intolerance.

Nearly every continent has recorded incidents. Much of the globe now communicates through social media, with approximately one-third of the worldwide people alone participating on Facebook social media. Individuals predisposed toward racism, sexism, or homophobia have discovered niches that might strengthen their ideas and provoke them to violence, according to specialists. Social media platforms also allow violent perpetrators to broadcast their actions. (Laub, 2019)

As raised in the literature review the same time that people are worried about hate speech also have the problem that people also are concerned about censorship and homogeneity of speech on social media. Analysing this research, we can observe that there were no participants that has consider free speech as not important. Therefore, both subjects can be considered equally important from a general point of view. The challenge arises however when freedom of speech can be transformed into hate speech and consequently when hate speech can be transformed into something else more serious. As stated in the literature review those stages can be developed very quickly inside the pyramid of hate. Laub, 2019 brings a graphic and some examples where hate crimes were committed due to the development of hate speech on social media.

Percent that agree "People should be able to make statements that are offensive to minority groups publicly" (2015)

Region	
United States	67
Latin America	50
Europe	46
Africa	36
Asia-Pacific	27
Middle East	24
Global median	35

Note: Displays the median among countries included in the survey.

Source: Pew Research Center.

COUNCIL on
FOREIGN
RELATIONS

(Source: Council on Foreign Relations, 2019)

- In Germany a correlation was found between anti-refugee Facebook posts by the far-right Alternative for Germany party and attacks on refugees. Scholars Karsten Muller and Carlo Schwarz observed that upticks in attacks, such as arson and assault, followed spikes in hate-mongering posts.

- In the United States, perpetrators of recent white supremacist attacks have circulated among racist communities online, and also embraced social media to publicize their acts. Prosecutors said the Charleston church shooter, who killed nine black clergy and worshippers in June 2015, engaged in a "self-learning process" online that led him to believe that the goal of white supremacy required violent action.

- In Myanmar, military leaders and Buddhist nationalists used social media to slur and demonize the Rohingya Muslim minority ahead of and during a campaign of ethnic cleansing. Though Rohingya comprised perhaps 2 percent of the population, ethnonationalists claimed that Rohingya would soon supplant the Buddhist majority. The UN fact-finding mission said, "Facebook has been a useful instrument for those seeking to spread hate, in a context where, for most users, Facebook is the Internet

- Sri Lanka has similarly seen vigilantism inspired by rumors spread online, targeting the Tamil Muslim minority. During a spate of violence in March 2018, the government blocked access to Facebook and WhatsApp, as well as the messaging app Viber, for a week, saying that Facebook had not been sufficiently responsive during the emergency.

And the examples are not limited to that, the concern is not only around hate speech by itself but also around what is the consequences that this behaviour can cause in society as a whole.

As raised in the literature review the same time that people are worried about hate speech also have the problem that people also are concerned about censorship and homogeneity

of speech on social media. Analysing this research, we can observe that there were no participants that has consider free speech as not important.

Free speech implies the liberty to speak in any way which does not infringe on the rights of others. You have the right to criticise the work of your national leaders for example. That's why free expression is so important in a democratic society. Democracy entails everybody in society making democratic decisions on the rules they live by and who enforces them. The open interchange of opinions, views, and information gives us the information we need to make such judgments. That is also why, under autocracies, free speech and the organs that sustain it, such as free media and civil society, are generally the first to go. (LibertiesEU, 2021)

Another challenge encountered concerning this subject is the fact that social media nowadays rely mainly on content review using artificial intelligence, algorithm systems, reporting tools designed by themselves and some people that are content moderators that work revising posts, however, it is possible to realise that the majority of alternatives are automated, despite the efforts human work in this matter face its challenges. Moderators, on the other hand, are overwhelmed by the enormous variety of data and the stress of filtering through distressing postings, and social media corporations do not allocate resources equitably across the various markets they cover. (Laub, 2019)

In addition to that as the robots do not understand the peculiarities and complexities of languages some are not illegal or not harmful content and users ended up being punished. Still, according to Laub, 2019 problems emerge when platforms' artificial intelligence is badly matched to local dialects and businesses have spent very few proficient people. He uses the example that what happened in Myanmar.

[…] Reuters reported, that Facebook employed just two Burmese speakers as of early 2015. After a series of anti-Muslim violence began in 2012, experts warned of the fertile environment ultranationalist Buddhist monks found on Facebook for disseminating hate speech to an audience newly connected to the internet after decades under a closed autocratic system.

In addition to that, it has the factor that currently social media are private companies where the monopoly is concentrated in just a few people's hands, which detain a certain power of manipulation and can shape important and historical events according to their interests. In conclusion, speech is not that easy to define, and there are more subjective and complex elements involved than just posts and statements done on social media.

5.3. Objective 3 – Explore the possibilities of using restorative practice to address hate speech.

The third and final objective of this research is to evaluate if restorative practice would be viable in addressing the hate speech issue on social media. According to the research was possible to realise that restorative practice is still not very popular nowadays, once the survey showed that most of the participants (59%) have never even heard about restorative practice and when asked what their knowledge about the subject a small percentage of them was only knowledgeable about (19%) or very knowledgeable about (5.7%).

Another point that this work raised was the issue the hate speech and the solution used until this moment is social media content moderation, social media tools and internal policies. Although some question arises "Is it effective?" or "what are the benefits?". Hassinof et al., 2020 make an interesting comparison, in their research, they compare the current systems used by social media to contain hate speech spread, which uses banishment of users or posts in a punitive way, with the traditional legal criminal system. According to them, victims' needs are frequently overlooked in both the criminal justice system and commercial content control. The individual who has been damaged has limited impact on the process of resolving the injury in the criminal justice system. Victims could be asked to testify, and provide an impact statement, to assist prosecutors in their efforts to convict, although they normally have very little say in the matter. They may be provided therapy and monetary recompense in some situations, including through victims' compensation programmes. Platforms have even fewer chances for victims to engage in a process or obtain defenders, assistance, or compensation than the criminal court system.

When comparing the two systems they indeed have a lot in common and taking into consideration that in social media content moderation leaves even smaller chances of the harmed person to have any participation another question emerged "Banishing people is it any beneficial?". Another research shows that the exclusion of a person or account from certain social media does not corroborate tackling hate speech. That occurs because they simply migrate from one place to another. If the social media has stricter policies and they happen to be banished from there they change for another social media that have more flexible policies or are more aligned with their opinions. About that Laub, 2019 illustrates with an example.

• The 2018 Pittsburgh synagogue shooter was a participant in the social media network Gab, whose lax rules have attracted extremists banned by larger platforms. There, he espoused the conspiracy that Jews sought to bring immigrants into the United States, and render whites a minority, before killing eleven worshippers at a refugee-themed Shabbat service. This "great replacement" trope, which was heard at the white supremacist rally in Charlottesville, Virginia, a year prior and originates with the French far-right, expresses demographic anxieties about nonwhite immigration and birth rates.

Likewise, AI algorithms that scan platforms for prohibited language and pictures create a veneer of cleanliness, relieving some of the popular pressure on platforms to tackle the concerns. However, the fundamental structures that allow individuals to harass continue to exist. From that, it is possible to realise that bashing and exclusion solutions only give the sensation of action. People think something is being done while it is just scratching the surface a bit more that it is possible to recognise that the problem was not solved.

Therefore it is possible to say that exclusion is not the best alternative to address this issue, considering that in most cases, the individual who was harassed is not usually contacted or informed about the process after making a report. In social media content moderation, the harmed person's best possible outcome is that bad information just vanishes. Typically, the harasser will cease since the platform complaint made them understand they had overstepped the limits. However, harassers frequently regard these repercussions to be unreasonable, unjust and punishing, and they either discover other methods to harass or finally quit out because they are bored or irritated and not because they learn something or they regret it. (Hasinoff et al., 2020)

In the survey people were asked if they have ever been a target of hate speech before when the participants answered yes they were led to another question that asked how was the social media support, this question has the aim to evaluate how was the participation of people that had suffered a hate speech attack "were they heard?", "Did someone from the social media talked to them?", "Did they have any advice from social media?" as can be noticed by the survey most of the participants (58.8%) did not have any support at all from social media.

On the other hand, the restorative practice has the aim to address the problem differently. Instead of only excluding the person that causes harm and presenting this as a punishment to the person that was harmed the approach is more humanistic and its techniques are focused on hearing what is the need of the person that has been harmed and bringing closer the aggressor to a transformative process not only to repair the harm but also to avoid a

reoccurrence. The practise is founded on the ideas of justice, respect, honesty, and collaborative problem-solving in the community. It involves Restorative Justice, which is a procedure in which violators and those who have been affected meet to deepen understanding and come up with cooperative and collaborative ways to rectify harm. (Project, n.d.)

To corroborate the fact that restorative practice is very effective in decreasing the chances of wrongdoing occurring the authors Hasinoff et al., 2020 state that restorative justice approaches have repeatedly shown value in a range of situations, especially in schools and workplaces, according to empirical studies. the harmed person and the aggressor are typically happier with the restorative process than with the traditional criminal legal system, and studies have found significant decreases in both violent and property crime repeat offending. Furthermore, the authors also reinforce that with examples.

[…] Some U.S. schools that have implemented restorative justice practices have benefited from improved school climate, dramatically decreased suspension and expulsion rates, and reductions in bullying.

Some schools have used in-person restorative justice practices to respond to and attempt to prevent online bullying. It's been used at the college level too. At Dalhousie University in Nova Scotia, Canada, male dentistry students created a private Facebook group that included sexist, misogynist, and homophobic remarks and images, as well as posts targeting some of their female classmates. Once these issues came to light in 2014, the school opted for a restorative justice process to address them. (Hasinoff et al., 2020)

In addition to that, the literature review has shown with the example of the Garda and the traveller family that restorative practice has a transformative effect when dealing with prejudice since the garda's actions were a result of his pre-concepts that he had about that community in specific. The restorative practice has changed that and has brought the participants together and more than that has transformed a person who before could keep spreading misinformation and prejudice into a person that respects and now help the community stop spreading hate.

Furthermore, some of the questions of the survey were designed to analyse if people would engage, consider or believe that this approach would be valid to address the hate speech issue. At the beginning of the survey was given a brief and general explanation about restorative practice as it was supposed that might have some people that would not know what the subject is about. The results have concluded that even though most of the people haven't even heard about the practice (59%) some of the people that answered yes believe that restorative practice could be effective in addressing hate speech in a total of 69.8%,

while 27.9% believe at least that the approach might be effective and only 2.3% disagree that this approach could work. Therefore all participants were asked if they would consider the approach a valuable way to cope with the problem and 42.2% replied yes, while the other 57.8% replied maybe and no answers were marked no. So with this perspective, it is possible to say that people are willing to try a new way to address the issues that hate speech brings to the social media sphere.

6. CONCLUSION

Hate speech is a new and delicate subject that has emerged quickly inside social media environments, which is challenging this generation to find an effective and consistent solution for the problem that it causes. From the findings of this research was possible to learn that hate speech has a systematic structure that's started based on biased attitudes, going through acts of bias, systemic discrimination, and bias-motivated violence is even possible to achieve a stage of genocide.

In the past, this was in a context where people were exposed or more exposed to this type of situation when they were in social places, nowadays with the internet and technological innovations online harm can be everywhere, and it is a concern how big it has grown and how much damage it can cause. Currently, people don't need to go out to suffer bullying, harassment or be a target of a hateful statement, they can be inside people's houses, in their work, on their hands with just a click and social media is the centrepiece of this phenomenon.

However, hate speech brings it is more complex. It is not as simple as separating the good from the evil, the black from the white, the things are more subjective and deeper, and with this comes the concern about free speech. As shown in this study previously free speech is a topic that is as much important as hate speech, since without free speech democracy is inexistent. In this study, it was possible to realise that social media has developed its way of trying to cope with the problem of hate speech, although this solution does not seem as effective as expected and this solution also comes with another challenge that is tackling hate speech respecting free speech. Nowadays the amount of content flow on the social media makes human content moderation not viable, and on the other hand, artificial intelligence and automated systems are not able to process complexities of the language and all of this ended up generating an exclusion of content that are not illegal or inappropriate, consequently affecting free speech.

Furthermore, social media has created a culture of banishment and informal punishment where people are simply excluded and don't even have a chance for change or regret. In this sense, it arises the doubt about the efficiency and benefits of the method. As this study has figured it out in an earlier stage this is a system that has almost no benefits for any of the parties. In those cases when an episode of hate speech happens the scenario usually is the same for everybody, the aggressor will be banished and can just migrate to another

social media and produce more hateful speeches, the content will simply disappear and the party that suffers the harm will probably even not know how that content just disappeared and what a happen with the aggressor. Consequently, we can conclude that systems used by social media currently just give us the impression that something is being done without being effective.

Contrastingly, restorative practice is an approach where people that have suffered harm are empowered by being able to participate in the process. In this model people's needs are heard and bring to the aggressor a sense of responsibility. Its name already suggests what is the focus of this approach, restorative comes from restoring, restoring the harm done and trying to get as closer as possible to the status quo of the person that was harmed. According to this work was possible to observe that restorative practice not only brings people together and decreases the chances of recurrence but also acts in a transformative way with the aggressor, changing the mentality from the prejudice and bringing even the possibility to transform that individual that was once an aggressor in an advocate for minorities and the communities harmed.

The benefits of this type of approach to address this type of harm are not limited to the people involved, it also extends to the social media interests and the social media environment as a whole. According to Hasinoff et al., 2020 as one of the social media aims is having long-term users it is recommended that social media companies invest in restorative and transformative justice techniques if they want to keep their users for the long run. Training and assisting community moderators in the application of these approaches might have long-term benefits such as reduced recurrence and an improved community atmosphere. In conclusion, the restorative practice has proven to be very beneficial in addressing hate speech, not only for the parties involved, but it is also beneficial in a wider sphere.

7. REFLECTION

If a person approached me three years ago and say that in three years, I would live an adventure of a lifetime I wouldn't have to believe it. If someone has said to me that I would move to another country on my own, I would have doubted it and if this person have said to me that I would be finishing a master's degree in another language, I would have doubted it even more.

There's no other word that describes this Master better than a challenge, a biter sweet challenge. In this course, we learned but more than that we cried, we celebrated, and we succeed. I remember all the nights spent awake because I was finishing an assignment, all the days that I thought of giving up, but as Brazilian music that I like says "I did come this far to give up now" and here I am now writing the final part of the most important assignment that I have ever done. This work changed me, not only all the learning that I got from this, and all the interesting discoveries that I have made, it makes me see people in a more humanistic way and showed that we are more profile pictures, likes and reactions and that sometimes we might forget but it has another human being behind the screens and keyboards, and if we look to that with a bit more attention we can realise that we could make the world a better place.

Speaking of human beings, it is not possible to finish this without mentioning and remembering all the amazing people that this course allowed me to meet. It was teachers that I will always admire and colleagues that I will carry in my heart for life, and even though our paths will follow different roads I will always be grateful for them have crossed my way and I will never forget those who lived this adventure with me.

This Master's was an adventure that taught me more than I imagine, I have learned new techniques and new theories, and I also learned a lot about myself, learned that doesn't matter how big and difficult the challenge seems to be with hard work and persistence I will overcome. This is just the end of this chapter that I close with a fulfilment sensation, and this is just the beginning of a new chapter.

"Know all the theories, master, all the techniques, but as you touch a human soul be just another human soul."

— *Carl Gustav Jung*

BIBLIOGRAPHY

Alkiviadou, N., 2021. *Human rights here.* [Online]
Available at: https://www.humanrightshere.com/post/automated-content-moderation-hate-speech-and-human-rights
[Accessed 26 April 2022].

Alter, C., 2022. *Time.* [Online]
Available at: https://time.com/6171183/elon-musk-free-speech-tech-bro/
[Accessed 30 April 2022].

Amy A. Hasinoff, A. D. G. N. S., 2020. *Tech Stream.* [Online]
Available at: https://www.brookings.edu/techstream/the-promise-of-restorative-justice-in-addressing-online-harm/
[Accessed 05 May 2021].

Basedo, 2020. *Basedo.* [Online]
Available at: https://besedo.com/resources/blog/what-is-content-moderation/
[Accessed 26 April 2022].

Beauchere, J., 2018. *Microsoft - Microsoft on the Issue.* [Online]
Available at: https://blogs.microsoft.com/on-the-issues/2018/10/03/millennials-encounter-the-most-risk-online-new-microsoft-research-shows/
[Accessed 06 March 2022].

Bimber, B. (2012). Digital media and citizenship. In Semetko, H., Scammel, M. (Eds.), The SAGE handbook of political communication (pp. 115–127). SAGE.

Bryman, A. (2012). Social research methods (5th ed.). Oxford: Oxford University Press.

change, R. S. f., n.d.. *RJ Strategies for change.* [Online]
Available at: https://restorativejustice.ie/hate-speech-restorative-conference/
[Accessed 28 April 2022].

council, R. J., n.d.. *Restorative Justice council.* [Online]
Available at: https://restorativejustice.org.uk/resources/evidence-supporting-use-restorative-justice#:~:text=The%20government's%20analysis%20of%20this,the%20Campbell%20Collaboration%20in%202013.
[Accessed 28 April 2022].

David Downes, P. R. C. C. C. G., 2007. *Crime, Soial control and Human Rights..* Cullompton, Devon: Willian Publishing.

David Phair, K. W., 2021. *Grad Coach.* [Online]
Available at: https://gradcoach.com/saunders-research-onion/
[Accessed 28 April 2022].

Dawkins, R., 2006. *The Selfish Gene.* 30th Anniversary edition ed. New York: Oxford University Press.

Emma Bell, A. B. B. H., 2022. *Business research methods.* Sixth edition ed. New York: Oxford University Press.

Emma Llansó, J. v. H. P. L. J. H., 2020. *Artificial Intelligence, Content Moderation, and Freedom of Expression,* s.l.: Transatlantic working group.

Gil de Zúñiga, H., Molyneux, L., Zheng, P. (2014). Social media, political expression, and political participation: Panel analysis of lagged and concurrent relationships. Journal of Communication, 64(4), 612–634.

Goddard, W. & Melville, S. (2004). Research Methodology: An Introduction, (2nd ed.) Oxford: Blackwell Publishing.

Golino, M. A., 2022. *Institute for Internet & Just Society.* [Online]
Available at: https://www.internetjustsociety.org/algorithms-in-social-media-platforms#:~:text=Algorithms%20in%20social%20media%20platforms%20can%20be%20defined%20as%20technical,actually%20engage%20with%20such%20content.
[Accessed 25 April 2022].

Gov.UK - Department for Digital, C. M. &. S., 2021. *Gov.UK.* [Online]
Available at: https://www.gov.uk/guidance/understanding-and-reporting-online-harms-on-your-online-platform
[Accessed 06 March 2022].

Irati Vizcarguenaga-Aguirre, J. R. L.-R., 2020. *Mono, mixed or multiple strategy approach: a descriptive study of the latest published articles in the International Journal of Project Management.* Spain, University of the Basque Country.

Ireland, N. Y. C. o., n.d.. *No Hate Speech.* [Online]
Available at: https://www.youth.ie/programmes/equality-intercultural/no-hate-speech/#:~:text=Hate%20speech%20covers%20all%20forms,that%20is%20based%20on%20intolerance.
[Accessed 16 Februrary 2022].

Johannes Baldauf, J. E. a. J. G., 2019. *Hate Speech and Radicalism Online,* Toronto: ISD.

Katie Moon, D. B., 2017. *Integration and Implementation Insights.* [Online]
Available at: https://i2insights.org/2017/05/02/philosophy-for-interdisciplinarity/
[Accessed 28 April 2022].

Kleinman, M., 2022. *Amnesty International UK.* [Online]
Available at: https://www.amnesty.org.uk/press-releases/tech-twitter-must-not-willfully-turn-blind-eye-violent-and-abusive-speech
[Accessed 28 April 2022].

Laub, Z., 2019. *Council on Foreign Relations.* [Online]
Available at: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons#chapter-title-0-4
[Accessed 11 March 2022].

Laub, Z., 2019. *Council on Foreign Relations.* [Online]
Available at: https://www.cfr.org/backgrounder/hate-speech-social-media-global-

comparisons
[Accessed 05 May 2022].

Latimer, J., Dowden, C., & Muise, D. (2005). The Effectiveness of Restorative Justice Practices: A MetaAnalysis. Prison Journal, 85(2), 127-144.

league, A.-d., 2021. *ADL.* [Online]
[Accessed 28 Februrary 2022].

LibertiesEU, 2021. *Liberties.* [Online]
Available at: https://www.liberties.eu/en/stories/free-speech-social-media/43773
[Accessed 05 May 2022].

Marcus, G. E., Sullivan, J. L., Theiss-Morse, E., & Wood, S. L. (1995). With Malice Toward Some: How People Make Civil Liberties Judgments. Cambridge: Cambridge University Press.

Mark N. K Saunder, P. L. A. T., 2019. *Research methods for business students.* Eighth ed. New York: Pearson.

Matelski, K., 2018. *Vizion.* [Online]
Available at: https://www.vizion.com/blog/what-is-a-social-media-algorithm-and-how-does-it-work/
[Accessed 24 April 2022].

Merriam-Webster, n.d.. *Merriam-Webster.* [Online]
Available at: https://www.merriam-webster.com/dictionary/harm
[Accessed 04 Februrary 2022].

Merriam-Webster, n.d.. *Merriam-Webster.* [Online]
Available at: https://www.merriam-webster.com/dictionary/viral
[Accessed 25 April 2021].

Oireachtas Library & Research Service, 2020, L&RS Note: Online harms – what are the legal issues?

Ortiz-Ospina, E., 2019. *Our World in Data.* [Online]
Available at: https://ourworldindata.org/rise-of-social-media
[Accessed 05 May 2022].

PDST, n.d.. *PDST.* [Online]
Available at: Restorative practice is founded largely on a set of basic ideals and the intentional cultivation and strengthening of certain skills such as empathy and problem-solving abilities. This approach enables the development of trust between and among individual
[Accessed 19 April 2022].

Perrigo, B., 2019. *Time.* [Online]
Available at: https://time.com/5739688/facebook-hate-speech-languages/
[Accessed 24 April 2022].

Philosophy, S. E. o., 2022. [Online]
Available at: https://plato.stanford.edu/entries/hate-speech/#Bib
[Accessed 16 Februrary 2022].

Project, L. R. P., n.d.. *Limerick Restorative Practice Project.* [Online]
Available at: http://www.restorativepracticeslimerick.ie/what-is-restorative-practices.html
[Accessed 07 May 2022].

QuestionPro, n.d.. *QuestionPro.* [Online]
Available at: https://www.questionpro.com/article/likert-scale-survey-questions.html
[Accessed 03 May 2020].

Reference, 2020. *Reference.* [Online]
Available at: https://www.reference.com/world-view/target-population-research-d8db8e47d6b12de5
[Accessed 01 May 2022].

Salehi, N., 2020. *Berkeley School of Information.* [Online]
Available at: https://www.ischool.berkeley.edu/news/2020/restorative-justice-approach-social-media-moderation
[Accessed 08 March 2022].

Service, O. L. &. R., n.d. s.l.: s.n.

SmartSurvey, n.d.. *SmartSurvey.* [Online]
Available at: https://www.smartsurvey.co.uk/survey-questions/likert-scale#:~:text=Likert%20Scale%20questions%20are%20a,or%20disagree%20with%20that%20statement.
[Accessed 02 May 2022].

Strategy, C. o. E. G. E., 2016. *Combating Sexist Hate Speech,* s.l.: Council of Europe.

SurveyMonkey, n.d.. *SurveyMonkey.* [Online]
Available at: https://www.surveymonkey.com/mp/gathering-demographic-information-from-surveys/
[Accessed 02 May 2022].

Thesismind, 2019. *Thesismind.* [Online]
Available at: https://thesismind.com/analysis-of-saunders-research-onion/
[Accessed 28 April 2022].

Thesismind, 2019. *Thesismind.* [Online]
Available at: https://thesismind.com/types-of-research-methods-or-methodology/
[Accessed 01 May 2022].

University, A., n.d.. *American University.* [Online]
Available at: https://programs.online.american.edu/msme/masters-in-measurement-and-evaluation/resources/emic-and-etic
[Accessed 28 April 2022].

University, M., n.d.. *Maryville University.* [Online]
Available at: https://online.maryville.edu/blog/evolution-social-

media/#:~:text=In%201987%2C%20the%20direct%20precursor,social%20media%20platf
orm%20was%20launched.
[Accessed 04 May 2022].

Windham, C., n.d.. *Azcentral.* [Online]
Available at: https://yourbusiness.azcentral.com/quantitative-qualitative-datacollection-
methods-13608.html
[Accessed 01 May 2022].

Zehr, H., 2015. *Changing Lenses.* Twenty-fifth anniversary edition ed. kitchener: Herald
Press.

APENDIX

A – Questionnaire

Restorative Practice as an alternative to addressing hate speech on social media

The purpose of this research is to analyse the impacts of hate speech spread on social media and
understand what strategies platforms are using to contain the problem and if those measures are
effective. Furthermore, the research will be valuable to assess how an alternative remedy such as restorative practice could be effective in combating the problem. In conclusion, the research will look at solutions and effective techniques for dealing with a global issue.

Restorative Practice is a method of dealing with conflict that focuses on repairing the harm done. It is a way of settling issues in which all parties involved can participate.

This survey is being conducted by the student Mariane N. G. Kobori under the supervision of professor Gareth Leech. And will be used as a part of a final project to conclude a Master's Degree in Dispute Resolution at Independent College Dublin and it will remain confidential and will be used only for this purpose.
To protect the confidentiality, this survey does not contain any personal information enabling us to identify who is answering. This survey is also voluntary any participant is free to withdraw the survey at any time.

By submitting your responses, you are agreeing to participate in the survey and allowing your answers to be used as part of this project.

1. 1. What is your gender?

Mark only one oval.

◯ Male

◯ Female

◯ Prefer not to say

◯ Other: _____

2. 2. How old are you?

Mark only one oval.

○ 18 - 24
○ 25 - 34
○ 35 - 44
○ +45

3. 3. Do you use social media often?

Mark only one oval.

○ Hourly
○ Daily
○ Weekly
○ Monthly
○ Never

4. 4. Which of the following social media do you use the most?

Mark only one oval.

○ Facebook
○ Instagram
○ TikTok
○ Snapchat
○ Other: _____

5. 5. How often do you witness sexists speech on social media that you use the most?

Mark only one oval.

- ⬭ Always
- ⬭ Often
- ⬭ Sometimes
- ⬭ Rarely
- ⬭ Never

6. 6. How often do you witness racist speeches on social media that you use the most?

Mark only one oval.

- ⬭ Always
- ⬭ Often
- ⬭ Sometimes
- ⬭ Rarely
- ⬭ Never

7. 7. How often do you witness religious intolerance speeches on social media that you use the most?

Mark only one oval.

- ⬭ Always
- ⬭ Often
- ⬭ Sometimes
- ⬭ Rarely
- ⬭ Never

8.   8. How often do you witness political intolerance speeches on you use the most?

Mark only one oval.

◯ Always

◯ Often

◯ Sometimes

◯ Rarely

◯ Never

9.   9. How often do you witness homophobic or LGBTQIA+ intolerance speeches on social media that you use the most?

Mark only one oval.

◯ Always

◯ Often

◯ Sometimes

◯ Rarely

◯ Never

10.   10. How likely are you to engage in social media against the speeches described above?

Mark only one oval.

◯ Definitely

◯ Probably

◯ Possibly

◯ Probably not

◯ Definitely not

11.   11. How important is discussing hate speech?

Mark only one oval.

◯ Very important

◯ Important

◯ Somewhat important

◯ Not important

12.   12. How important is free speech?

Mark only one oval.

◯ Very important

◯ Important

◯ Somewhat important

◯ Not important

13.   13. Have you ever been a target of hate speech on social media?

Mark only one oval.

◯ Yes

◯ No   *Skip to question 16*

14.   14. If yes. Did this event affect you?

Mark only one oval.

◯ Not at all

◯ Very little

◯ Fairly well

◯ Quite well

◯ Very well

15.   15. Did you have any support from the social media
        community?

        Mark only one oval.

        ◯ None
        ◯ Slightly
        ◯ Considerable
        ◯ Great

16.   16. Did you ever use any tools provided by social media to
        report hate speech?

        Mark only one oval.

        ◯ Yes
        ◯ No

17.   17. What is the importance of having tools and internal
        measures inside social media to cope with hate speech?

        Mark only one oval.

        ◯ Extremely Important
        ◯
        ◯