



D2.1 – Consolidated knowledge and best practices



HATEDEMICS is funded by the European Union
(CERV-2023-CHAR-LITI-101143249)

Project Information

Grant Agreement	101143249
Name	Hampering hate speech and disinformation through AI-based technologies to prevent and combat polarisation and the spread of racist, xenophobic and intolerant speech and conspiracy theories
Acronym	HATEDEMICS
Call	CERV-2023-CHAR-LITI-SPEECH
Topic	Topic 4 - Protecting EU values and rights by combating hate speech and hate crime
Consortium Coordinator	Fondazione Bruno Kessler (FBK)
Project duration	24 months
Project start date	1 April 2024
Project closing date	31 March 2026
Coordinator	Marco Guerini

Document Information

Workpackage	WP2
Deliverable	D2.1
Deliverable Lead	FUNDEA
Dissemination Level	PUBLIC
Type	REPORT
Date due for submission	30/09/2024
Date submitted	17/10/2024
Authors	Diego de Haro (CENTRA) Rubén Martín (CENTRA) Eva Cataño (CENTRA) Lucía García del Moral (FUNDEA) Kinga Klich (DEMAGOG) Aleksy Szymkiewicz (DEMAGOG) Anna Kolos (NASK) Agnieszka Karlińska (NASK) Jan Dąbkowski (CEO) Maryna Manchenko (CESIE) Rachael Scicluna (SOS MALTA)
Editor(s)	Lucía García del Moral Martín (FUNDEA)
Reviewers	Maryna Manchenko (CESIE) Rachael Scicluna (SOS MALTA)



Revision History

Version	Date	Author	Comments
V0.1	10/10/2024	Lucía García del Moral Martín (FUNDEA)	First draft of the deliverable
V0.2	14/10/2024	Maryna Mancheko (CESIE)	Reviewer's evaluation
V0.3	15/10/2024	Rachael Scicluna (SOS Malta)	Reviewer's evaluation
V1.0	17/10/2024	Lucía García del Moral Martín (FUNDEA)	Final version of the deliverable

Copyright © 2024 HATEDEMICS Consortium Partners. All rights reserved. HATEDEMICS is a Citizens, Equality, Rights and Values Programme (CERV) project supported by the European Union under Grant Agreement (GA) no. 101143249. You are permitted to copy and distribute verbatim copies of this document, containing this copyright notice, but modifying this document is not allowed.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or EACEA. Neither the European Union nor the granting authority can be held responsible for them.



EXECUTIVE SUMMARY

The HATEDEMICS project, funded by the European Union's Citizens, Equality, Rights, and Values Programme (CERV), aims to tackle the rising issues of hate speech and disinformation online, which undermine social cohesion and fuel polarization. This report, part of Work Package 2 (WP2), develops a theoretical framework utilizing interdisciplinary theories like Social Identity Theory and Framing Theory to understand how these phenomena spread in digital spaces.

The report emphasizes the need for media literacy and digital resilience, equipping citizens with the skills to critically evaluate online content. It also addresses the legal challenges faced by EU member states in combating hate speech and disinformation, advocating for harmonized policies to enhance protection and accountability.

Findings from interviews and focus groups indicate that hate speech is increasingly normalized in digital environments, often targeting marginalized groups. Additionally, social media algorithms create echo chambers that amplify extremist views.

To address these challenges, the report highlights the importance of educational initiatives and AI-based tools aimed at promoting inclusivity and countering hate speech within the EU.



Table of Contents

EXECUTIVE SUMMARY	4
Acronyms	6
1 INTRODUCTION	7
2 KNOWLEDGE ON THE CHALLENGES OF HATE SPEECH and DISINFORMATION	8
2.1 Contemporary theories on hate speech and disinformation	8
2.1.1 Challenges hate speech and disinformation	8
2.1.2 Conspiracy theories in partner countries	18
2.2 Hate speech in the EU context	25
2.2.1 The state of implementation of the relevant EU regulations and policies	25
2.2.2 General overview in partner countries	31
3 QUALITATIVE and QUANTITATIVE ANALYSIS OF EXISTING BIG DATA.....	36
4 THE STRATEGIES AND GOOD PRACTICES TO FACE HATE SPEECH AND DISINFORMATION	38
5 CONCLUSION	40
6 RECOMMENDATIONS.....	41
REFERENCES	43



Acronyms

ABBREVIATED	EXTENDED
AI	Artificial Intelligence
CERV	Citizens, Equality, Rights, and Values Programme (European Union)
FRA	European Union Agency for Fundamental Rights
ICCPR	International Covenant on Civil and Political Rights
ILGA	International Lesbian, Gay, Bisexual, Trans and Intersex Association
LGTBIQ+	Lesbian, Gay, Bisexual, Transgender, Intersex, Queer, and other identities.
MENA	Spanish controversial term for Unaccompanied Foreign Minors
ODIHR	Office for Democratic Institutions and Human Rights
OSCE	Organization for Security and Cooperation in Europe
TFEU	Treaty on the Functioning of the European Union
UN	United Nations
WP	Work Package



1 INTRODUCTION

The HATEDEMICS project, funded by the European Union's Citizens, Equality, Rights, and Values Programme (CERV), aims to address the growing challenges posed by the spread of hate speech and disinformation, particularly within online environments. These phenomena have a profound impact on social cohesion, exacerbating polarization and fostering the rise of racist, xenophobic, and intolerant narratives across Europe. Hate speech, which targets individuals or groups based on characteristics such as race, religion, ethnicity, or gender, is further amplified by disinformation—deliberately false information designed to mislead and manipulate public opinion.

This report, as part of Work Package 2 (WP2), offers a comprehensive desk research analysis to explore the theoretical frameworks that explain the mechanisms behind hate speech and disinformation. This report, a key deliverable within Work Package 2 (WP2), presents a comprehensive desk research analysis aimed at developing a theoretical framework to better understand the underlying mechanisms that drive the spread of hate speech and disinformation. Through an interdisciplinary approach, theories such as Social Identity Theory, Framing Theory, and the Spiral of Silence are used to provide insights into how hate and false narratives spread and why they persist, especially in digital spaces where anonymity and echo chambers enhance their influence. These theoretical perspectives are crucial for understanding the dynamics of hate speech, which often emerges in polarized environments where individuals seek to reinforce in-group identities while dehumanizing out-groups.

In addition to these psychological and sociological theories, this report addresses the growing importance of media literacy and digital resilience in combating the effects of hate speech and disinformation. As online platforms become central to public discourse, the ability to critically assess media content and recognize manipulative or false information has become an essential skill for modern citizens. Educational initiatives that promote critical thinking and empower individuals to challenge harmful narratives are crucial to building more resilient societies capable of resisting both disinformation and hateful rhetoric.

Furthermore, this deliverable highlights the legal and regulatory challenges faced by European Union member states in combating hate speech and disinformation. While the EU has implemented several key directives and frameworks, including the 2008 Framework Decision on combating racism and xenophobia, the application and enforcement of these laws vary significantly across member states. The lack of a harmonized legal approach, especially in relation to emerging digital environments, creates gaps in protection and accountability. This report underscores the need for more robust and coordinated EU-wide policies that address the complex intersection between hate speech, disinformation, and digital technology.

The deliverable also incorporates findings from national interviews and focus groups conducted across several EU member states. These qualitative investigations reveal key common trends regarding the prevalence of hate speech and disinformation. Participants across different countries consistently reported that hate speech has become normalized in digital spaces, with some groups, such as migrants and LGBTIQ+ individuals, being frequent targets. Moreover, disinformation is often intertwined with hate speech, as false narratives—



ranging from conspiracy theories to misleading political rhetoric—fuel divisive attitudes and hostility towards marginalized groups.

Another major theme that emerged from the interviews and focus groups is the growing role of social media in spreading hate speech. Algorithms that promote content engagement often contribute to the creation of echo chambers, where individuals are exposed to more extreme and inflammatory rhetoric. The anonymity of online platforms also emboldens users to express hateful opinions without fear of repercussions, further entrenching polarized views.

Addressing these challenges requires not only legal frameworks but also educational interventions. The report emphasizes the need to enhance media literacy to equip individuals with the critical thinking skills necessary to recognize and resist disinformation and hate speech. Participants also stressed the importance of promoting digital resilience among the general population, helping individuals to navigate online spaces responsibly and thoughtfully.

This report, therefore, lays the groundwork for understanding the complex interactions between hate speech, disinformation, and social dynamics in digital environments. By integrating both theoretical and empirical insights, it aims to inform the development of effective AI-based tools and educational strategies to combat hate speech and foster a more inclusive, tolerant society within the European Union.

2 KNOWLEDGE ON THE CHALLENGES OF HATE SPEECH and DISINFORMATION

2.1 Contemporary theories on hate speech and disinformation

2.1.1 Challenges hate speech and disinformation

The challenges on hate speech and disinformation have major effects on the functioning of political systems and the resilience of democratic societies. Over recent decades, more studies elucidate the societal contexts in which hate speech and disinformation thrive, each offering insights into the mechanisms that facilitate their proliferation. Collectively, these theoretical frameworks enhance our understanding of the operational dynamics of hate speech and disinformation in contemporary Western societies, while also illuminating the challenges these phenomena pose to social cohesion and democratic discourse, but especially on the daily experiences of individuals and those professionals supporting victims of hate speech and crime.

Recent research based on interdisciplinary approaches has focused on seven dimensions to address contemporary theories on hate speech and disinformation. Given these considerations, this deliverable within the Hatedemics project (WP2) attempts to classify a theoretical framework in the following main areas of study.

Social Identity Theory (SIT)



Social Identity Theory posits that individuals categorize themselves and others into distinct groups, often leading to in-group favouritism and out-group hostility. This dynamic can foster the emergence of hate speech, as individuals may dehumanize those outside their group, thereby justifying discriminatory and aggressive behaviours.

Social Identity Theory, developed by Henri Tajfel and John Turner in the 1970s introduces social identity as a way of understanding intergroup behaviour (1979, 1982). His work focused on how individuals categorize themselves and others into groups, and how this influences behaviour. According to this work, individuals derive a significant portion of their identity from their membership in social groups. The SIT embeds four main concepts to understand social behaviour: categorization, in-group favouritism, out-group hostility and social comparison.

First, individuals categorize themselves and others into groups (in-groups and out-groups) based on various characteristics such as race, religion, or social affiliation. This cognitive process aids individuals in navigating their social environment. Second, members of an in-group frequently exhibit preferential treatment toward their own group, resulting in positive bias toward in-group members and negative bias toward out-group members. This phenomenon can manifest as favouritism, solidarity, or cooperation within the in-group. Third, the perception of out-groups often leads to prejudice and discrimination. Individuals may dehumanize those outside their group (out-group hostility), thereby justifying negative behaviours, including hate speech and aggression (Turner et al. 1987). Lastly, individuals engage in social comparisons to enhance their self-esteem. Favourable comparisons with out-groups can bolster in-group identity, while unfavourable comparisons may provoke a sense of threat and hostility toward out-group members.

Thus, hate speech is often directed towards out-groups as a way to reinforce the superiority of the in-group. By engaging in hate speech, individuals signal their loyalty to their group and devalue the out-group to increase their own self-esteem or sense of group cohesion (Reicher et al. 2008). Their behaviour is oriented to show their affiliation/membership of the group, the stability of the group, but also a sense of heroic behaviour from those leading on a specific issue. When the in-group feels threatened or perceives itself as inferior, individuals may use derogatory language or hate speech to devalue out-groups and elevate their own group. Research has also shown how the social and psychological factors interact to create hate dynamics, including group identity and intergroup conflict, that contribute to the perpetration of hate crimes, including hate speech (Perry, 2001). Hate speech often includes language that dehumanizes out-groups, portraying them as less than human, dangerous, or unworthy of rights (Esses et al. 2013).

Particularly, in online environments or closed communities where extreme ideas circulate, group polarization can lead to a radicalization of hate speech. As group members reinforce each other's negative views of the out-group, they may escalate from mild stereotyping to overt expressions of hatred and violence (Gagliardone et al. 2015). Other authors (Müller and Schwarz, 2020) look also at the relationship between online hate speech, group polarization, and real-world hate crimes, showing how group dynamics on social media can escalate hate.

Framing Theory



Framing Theory highlights the significance of how information is presented and interpreted. Frames are the perspectives or angles from which information is presented. They define what is important and how an issue is understood. Goffman (1974) is considered one of the founders of framing theory. His work in sociology focused on how individuals organize experiences and events using mental structures or frames. He suggested that people interpret social events based on the “frame” they apply to the situation. The process by which media or communicators construct and promote a particular frame for interpreting an event, issue, or situation (Entman, 1993). Moreover, framing theory is closely linked to the agenda-setting function of media, which focuses on how media can influence what issues people think about and how to blame individuals for societal problems, while thematic frames encourage people to think of issues as systemic or structural (Iyengar, 1991; Scheufele, 1999).

Hate speech can be strategically framed to normalize prejudiced views or incite violence, while disinformation campaigns manipulate perceptions and narratives, rendering harmful messages more palatable and believable. Related to hate speech, different frames attribute the causes to various factors, such as individual prejudice, societal conditions, or media influence (Roth, 2012). In this regard, the academic literature distinguishes between individual blame frame, and structural frame. Framing hate speech as resulting from individual prejudice or ignorance leads to an understanding that individuals are responsible for their actions. This might lead to calls for education and awareness programs to combat personal biases. Conversely, framing hate speech as a product of systemic inequality or socio-economic conditions emphasizes that addressing the issue requires changes in broader societal structures, such as promoting diversity and inclusion in educational institutions.

However framing theory can be applied to different subfields more than communication studies. From the psychosocial approach (Bandura, 1999; Loughnan et al. 2010; Perryman, 2001; Cohen, 2001; Haslam and Loughnan, 2014; Zick, 2018), frames can shape perceptions of victims and perpetrators of hate speech. On the one hand, framing individuals targeted by hate speech as victims emphasizes their suffering and the need for empathy and support. This can mobilize public opinion in favour of anti-hate speech legislation and victim support services (victimization frame). On the other hand, hate speech may also be framed in a way that dehumanizes the targets, portraying them as threats or undeserving of rights. This framing can perpetuate cycles of violence and discrimination, as it justifies the use of hate speech against certain groups (dehumanization frame). Some of these studies incorporate a center-victim approach (Bruneau and Saxe, 2010) emphasizing the importance of humanization in mitigating dehumanization and promoting empathy for victims. Other research also links frame theory to social identity theory (Dehue et al. 2014), analysing how group identity influences perceptions of victimization and the tendency to dehumanize others.

The frame theory is also used in hate speech legislation context. Framing of laws and policies regarding hate speech can significantly influence public opinion and legal outcomes. Following this approach, two main legal frames can be differentiated. On one side, protective frame, and on the other, free speech frame, giving birth to the debate of framing hate speech laws as limitations on free speech and expression.

The Spiral of Silence



This theoretical framework suggests that individuals may refrain from expressing their opinions if they perceive themselves to be in the minority. Valkenburg and Peter (2011) discusses the implications of online communication for the Spiral of Silence Theory, examining how online interactions can both mitigate and exacerbate the silencing of minority opinions. This phenomenon can create a chilling effect wherein hate speech silences dissenting voices and perpetuates harmful ideologies. Anonymity and audience along with disinformation exacerbates this issue by distorting perceptions of the prevalence of certain beliefs (Glynn et al. 2004). If audience perceive that hate speech is socially accepted or that there is a majority that supports it, they may be less likely to express dissenting views or advocate for stricter regulations against hate speech. The fear of being marginalized or ostracized can also silence individuals who want to express concerns about hate speech, particularly in environments where such views are seen as unpopular or radical.

Networked Publics and Echo Chambers

Networked Publics and Echo Chambers have emerged as a result of social media's rise, transforming the dissemination of hate speech and disinformation. The seminal paper of Noelle-Neumann (1974) established the basis for understanding how public opinion affects individual willingness to express dissenting views, including those related to hate speech.

Online platforms often facilitate the creation of echo chambers, where like-minded individuals reinforce each other's views, thereby complicating efforts to challenge extremist rhetoric and false information. Individuals who hold minority opinions about hate speech may feel isolated and refrain from expressing their views, reinforcing the dominance of pro-hate speech or indifferent stances in those online spaces (Boulianne, 2015). Cyberspace generates stimuli for polarisation and radicalization. Not only does it have the capacity to disseminate extremist messages massively, but they also use algorithms that create echo chambers, where like-minded individuals group together and reinforce each other's convictions.

In this regard, the participants in the national focus groups agree that there is a polarisation loop: the more content you receive from the topic, the more you will find content on networks and in the end, there is more polarisation because you consume that content massively. The nature of cyberspace therefore facilitates the legitimisation of extremist discourses related to several groups, as mentioned by the Italian report, such as women (also in politics), LGBTQ+ individuals, and ethnic or religious minorities, among others. According to the Maltese report, a greater differentiation in the degree of hate speech was observed between the different groups concerned. One key expert noted that women and ethnic minorities often face such speech on a daily basis, both in their personal lives and in the work environment. Some previous research applies this theory to perceptions of majority opinion influence discussions on sensitive topics, such as same-sex marriage, which parallels discussions on hate speech (Wang and Zhao (2016). Also other groups were identified as targets of multiple forms of hate simultaneously. For example, a migrant who identifies as LGBTIQ+ receives hate speech from several fronts.

When individuals from minority groups do speak out against hate speech, they may still find themselves silenced by the dominant narrative, leading to a suppression of their voices and experience (Wright et al. 2016; Schulz et al., 2016)). Victims of hate speech may feel particularly vulnerable and may choose to remain silent about their experiences for fear of



not being believed or facing further victimization. This silence can perpetuate the cycle of dehumanization and marginalization.

As these discourses reach a wide and varied audience, they begin to infiltrate public debate, intensifying clashes between different sectors of society. This phenomenon creates a vicious cycle in which disinformation feeds polarisation, and a polarised society becomes more vulnerable to disinformation. Therefore, cyberhate present in social networks, news forums, and the Internet is increasingly worrying due to its extreme nature and the freedom with which people express hate. This is highlighted by another participant from a more social perspective, proposing the responsible use of networks and instead of using that content to radicalise, use it to teach and deconstruct.

Cognitive Dissonance Theory

This theory developed by Festinger (1957) posits that individuals experience discomfort when confronted with information that contradicts their existing beliefs. It refers to the mental discomfort experienced when a person holds two or more contradictory beliefs, values, or attitudes, especially when they engage in behaviour that conflicts with their beliefs. This discomfort often leads individuals to change their beliefs or rationalize their behaviour to restore harmony.

In this context, hate speech and disinformation can reinforce pre-existing biases, leading individuals to reject opposing viewpoints and further entrenching their positions. Some authors (Cohen et al. 1994; Hollander, 2001; Willer and Fisher, 2017) have discussed how individuals rationalize their inaction against hate and discrimination due to social pressures while others (Latané and Darley, 1970) explore the cognitive dissonance experienced by individuals who witness hate or violence but choose not to intervene. In the specific context of hate crime, Perry (2001) analyses cognitive dissonance experienced by individuals who perpetuate or condone hate speech while holding conflicting values regarding equality and justice. In the same vein, Quinn and Lippman (1995) has previously studied how individuals involved in hate crimes reconcile their actions with their self-concept and social norms.

Critical Race Theory

Critical Race Theory examines the intersection of race and power within society, emphasizing that hate speech disproportionately impacts marginalized groups. This framework critiques systemic inequalities and elucidates how disinformation perpetuates racial stereotypes and injustices (Richardson and Colombo, 2014). Racial hate speech, in particular, serves to reinforce the subordination of racial minorities by normalizing discriminatory attitudes and behaviours. Scholars argue that hate speech is often protected by free speech laws, which fail to account for the historical and structural inequalities that such speech perpetuates (Brown, 2017; Delgado and Stefanic, 2017). In this sense, hate speech is seen not merely as an expression of individual prejudice but as a systemic issue rooted in racial power imbalances (Bell, 1992).

Academics also focus on counter-speech as a form of resistance to hate speech (Delgado, 1989). Conversely, racialised academics and communities often use alternative narratives instead of countering the narratives that hate speech promotes, creating spaces for empowerment and solidarity (Ladson-Billings, 1998; Solórzano and Yosso, 2002; Love, 2019).



In addition, both scholars and communities develop its work according to intersectional theory in racialized contexts (Crenshaw, 1995).

The normalisation of hate speech was a recurring theme in the Maltese report, with race as a common denominator. Many noted how social media is rife with hate speech towards migrants on a daily basis. One participant commented that '[hate speech] has been normalised, not only by the public, but also by the victims themselves'. One NGO representative explained that this situation arises in most meetings with their clients, adding: "It's so prevalent that they just downplay it, saying they have bigger problems right now".

Post-Truth Politics

It describes a political landscape in which emotional appeal and personal beliefs overshadow objective facts so that emotion, identity and personal belief has significant implications for the proliferation of hate speech (Suiter, 2016). In post-truth politics, political discourse often prioritizes emotion and identity over verifiable facts. In this vein, some authors provide a comprehensive examination of the post-truth phenomenon and its implications for democratic societies, including the rise of hate speech fuelled by emotional and ideological appeals rather than facts (McIntyre, 2018). Hate speech thrives in this environment because it appeals to people's fears, global narratives of persecution and hate (and murders), insecurities, existential experiences, and anger. False narratives about marginalized groups (migrants, racial minorities, religious groups) gain traction because they resonate emotionally, even when they are demonstrably untrue.

Post-truth politics often involves the deliberate spread of misinformation (false information shared unintentionally) and disinformation (false information shared deliberately) to manipulate public opinion (Marwick and Lewis, 2017; Wardle and Derakhsab, 2017; Lewandowsky et al. 2017). Hate speech can be bolstered by disinformation campaigns that target specific groups with lies and inflammatory rhetoric. Also, the rise of fake news in post-truth politics contributes directly to the spread of hate speech (Sunstein, 2009). Fake news articles and memes that promote racist, xenophobic, or homophobic messages gain viral traction, often without any factual basis and more importantly travel beyond borders.

Social media platforms, which play a central role in post-truth politics, allow disinformation to spread quickly. Hate groups and political actors can use these platforms to spread divisive messages about minorities, often with little regard for truth or accountability. The internet, particularly social media, creates echo chambers and filter bubbles where users are exposed primarily to content that aligns with their existing beliefs. In post-truth politics, these spaces amplify hate speech because users are less likely to encounter alternative perspectives or fact-based counterarguments. When people continuously consume content that vilifies certain groups, without encountering factual corrections or alternative viewpoints, their biases can become more extreme, leading to the normalization of hate speech.

In such an environment, hate speech and disinformation flourish, as emotional resonance often eclipses factual accuracy. This framework is ripe for hate speech, as it encourages a paradox old view where certain groups are considered threats to society. Hate speech can target minorities, migrants, or marginalized groups by portraying them as dangerous "others" who threaten the identity, security, or values of the dominant group. For example, conspiracy



theories collected in the WP2 national activities point out the spread of conspiracy theories about migrants in the four pilot countries (see the following section). These emotionally charged falsehoods contribute to a climate of fear and resentment, which can manifest in hate speech both online and offline.

Regulatory Approaches

The contemporary theories around hate crime and speech also explore the effectiveness of various legal and regulatory responses to hate speech and disinformation. These approaches range from advocating for stricter laws to promote accountability to calls for self-regulation by social media platforms.

The absence of a harmonized definition of hate crime and hate speech poses significant challenges across legal, social, and political domains. The differences in the Member States legal frameworks constitute a challenge to the EU regulation, as different jurisdictions define hate crimes and hate speech based on diverse cultural, historical, and legal contexts, leading to inconsistencies in the reporting, investigating, prosecuting, and preventing hate crimes. Applied research has identified a lack of clear definitions of hate crime, structured frameworks on hate crime reporting, recording, data sharing and resources between different EU countries. Furthermore, the criteria for categorizing hate crimes often differ, with jurisdictions unable to agree on the specific characteristics such as race, ethnicity, religion, sexual orientation, or disability. This discrepancy can result in gaps in legal response and victim's protection.

Internationally, human rights instruments such as the International Covenant on Civil and Political Rights (ICCPR) offer guidance, yet member states frequently interpret these provisions variably, resulting in a patchwork of protections that may inadequately address the needs of individuals facing hate-based violence or discrimination. Interpretations can be heavily influenced by prevailing political and social climates. Consequently, actions considered as hate speech in one context may be viewed as a legitimate expression in another, creating potential for censorship and legal disputes. This ambiguity can hinder effective law enforcement and diminish victims' trust in the justice system, discouraging them from reporting incidents.

Balancing the imperative of safeguarding free speech with the necessity of preventing harm poses an ongoing challenge; overly broad definitions risk infringing on legitimate expression, while overly narrow definitions may fail to protect vulnerable populations adequately. Moreover, cultural sensitivity is crucial, as terms and concepts related to hate can differ significantly across cultures, complicating the development of universally applicable definitions. Ultimately, enhancing public awareness and education about hate crimes and hate speech can facilitate a more nuanced understanding and foster effective responses to these pressing societal issues.

Thus, the lack of a harmonized definition of hate crime and hate speech poses significant challenges across legal, social, and political contexts (see the following sections). The main challenges can be summarised in diverse legal frameworks, varying criteria for hate crimes, subjectivity in legal and political interpretation of hate speech. Calls for harmonization are prevalent among advocacy groups and policymakers, as clearer definitions and guidelines are



essential for effective legal protections and international cooperation. Without a standardized definition, law enforcement and public agencies may struggle to consistently identify, categorize, and respond to hate crimes and speech. This inconsistency can undermine victims' trust in the legal system and discourage reporting. This can exacerbate feelings of vulnerability and a gap in the victim's protection, as well as discrepancies in the EU country's protection.

In summary, the lack of a harmonized definition of hate crime and hate speech presents complex challenges that affect legal, social, and cultural dynamics. Establishing clearer, more inclusive definitions could help improve protections and responses to these issues globally. The definition of hate speech is articulated through various perspectives from partner country. Interviewees highlight that hate speech promotes intolerance towards minorities, while another adopts the Council of Europe's definition, which includes hate speech as speech that incites exclusion or discrimination in categories such as ethnicity, race, gender and disability.

In the Italian report, participants agree in describing hate speech as expressions that incite hatred or violence, focusing on marginalised groups. One interviewee highlights that hate speech promotes intolerance towards minorities, while another adopts the Council of Europe's definition, which includes hate speech as speech that incites exclusion or discrimination in categories such as ethnicity, race, gender and disability. Emphasis is also placed on the need to differentiate between hate speech and other forms of verbal violence, such as bullying. Other contributors suggest that misinformation or misleading information can be catalysts for online hate speech, noting that while they are not always the direct cause, they are often connected, especially in racist, misogynistic or anti-LGBTQ+ narratives.

The Malta report also notes that all experts and practitioners call for an understanding of hate speech as diffuse and elusive, and often difficult to determine, especially when actions and linguistic comments are implicit and unclear, including non-inclusive language and non-verbal behaviour that is difficult to prove as abusive. They therefore understand hate speech as a fluid, ambiguous and difficult to delimit category, with multiple facets that depend on the social, political, economic and cultural context. Disentangling the various elements that make up the framework of hate speech included factors such as intent to cause harm, racist overtones both linguistic and cultural, non-verbal or passive behaviours, lack of critical and empathetic thinking skills, and the political and economic situation in Malta.

The Polish report aligns itself with the notion that hate speech is a violation of human rights, defining it as an expression whose purpose is to cause harm to others. This approach reinforces the idea that countering hate speech involves preventing harm. Several interviews, particularly those with lawyers and practitioners from international organisations, refer to the Council of Europe definition and the provisions of the Polish Penal Code. One interviewee highlights that hate speech, especially that related to misinformation, affects not only those who generate it, but also those who consume it, including young people, who may normalise this type of communication after years of exposure on social media. This highlights the responsibility of institutions to protect the right of citizens to participate in both physical and virtual spaces without being exposed to violence.



While the Polish aligns itself with the notion that hate speech is a violation of human rights, defining it as an expression whose purpose is to cause harm to others. This approach reinforces the idea that countering hate speech involves preventing harm. Several interviews, particularly those with lawyers and practitioners from international organisations, refer to the Council of Europe definition and the provisions of the Polish Penal Code. One interviewee highlights that hate speech, especially that related to misinformation, affects not only those who generate it, but also those who consume it, including young people, who may normalise this type of communication after years of exposure on social media. This highlights the responsibility of institutions to protect the right of citizens to participate in both physical and virtual spaces without being exposed to violence.

However, in the case of Malta and Spain, the experts and practitioners stated that hate speech can be difficult to capture because it is diffuse and elusive, and often difficult to determine, especially when actions and linguistic comments are implicit and unclear, including non-inclusive language and non-verbal behaviour that is difficult to prove as abusive. They therefore understand hate speech as a fluid, ambiguous and difficult to delimit category, with multiple facets that depend on the social, political, economic and cultural context. Disentangling the various elements that make up the framework of hate speech included factors such as intent to cause harm, racist overtones both linguistic and cultural, non-verbal or passive behaviours, lack of critical and empathetic thinking skills, and the political and economic situation.

Psychological Theories of Aggression

It investigates the psychological factors that drive individuals to engage in hate speech, including frustration, social learning, and deindividuation in online contexts. Apart from seminal works from psychology, Berkowitz (1990, 1993) expanded on the Frustration-Aggression Hypothesis and proposed the Cognitive Neoassociation Model, which posits that aggression is triggered not only by frustration but by negative emotions such as anger or discomfort. Aggression arises when these negative emotions lead to aggressive thoughts and behaviours. Other authors (Zillman, 1983, 1994) suggests that arousal from one situation can spill over into subsequent situations, increasing the likelihood of aggression. If someone is physically or emotionally aroused (e.g., after exercising or being frightened), they may misattribute this heightened arousal to a new situation, leading to more aggressive reactions. Later research (Anderson and Bushman, 2001; Anderson, 2002) integrates a general model based on various factors such as personality, situational variables, and cognitive processes to explain aggression. Also, deindividuation theory (Postmes and Spears, 1998) refers to a psychological state where individuals lose their sense of personal identity and self-awareness, leading them to engage in impulsive and often aggressive behaviour. This often occurs in group settings where anonymity is high. People in deindividuated states are more likely to act aggressively because they feel less personally accountable for their actions. Lastly, this approach was mixed with the social identity theory and cognitive dissonance theory previously explained.

Regarding hate crimes, the frustration-aggression hypothesis is often used for explaining emergence of hate speech when individuals or groups feel frustrated by their circumstances (e.g., economic hardships, social displacement) (Matsumoto, 2020).



Cumulative Radicalisation

All of the interviewees from the Italian report agree that conspiracy theories contribute to misinformation and disinformation and acknowledge their dangerous potential, considering antisemitism and racism of all types as the most common undercurrents in conspiracy theories. This could be explained through three common elements between hate speech and violent radicalisation: the existence of a community perceived to be under imminent threat, the identification of an external group as the source of that threat, and a sense of extreme urgency to act. As an example, the so-called Great Replacement conspiracy theory, which warns of the extinction of the white race in Europe and the United States due to the significant influx of immigrants from Africa and the Arab world and feminist/queer policies. For instance, in Malta this was expressed by the feeling that a population or a person is losing control of the situation through an unprecedented demographic shift and population densification.

Hence, in the three countries we can observe similar behavioural traits with cultural specificities.

Violent acts or extreme rhetoric can generate a backlash from the aggrieved group, triggering a spiral of escalation in which various forms of extremism feed off each other and perpetuate conflict. In Malta, (like in other countries if this is relevant) this type extremism appeared to be cyclical based on what the country is going through. For example, during Pride week a factchecker stated that there are more conspiracy theories around LGBTIQ+ people being associated with grooming and paedophilia. This was also supported by an NGO which stated that such extreme remarks are becoming common and certain public figures are the initiators of such false claims and misinformation about what queer means.

The above phenomena, known as reciprocal radicalisation or cumulative extremism, describe how different forms of extremism interact and reinforce each other, often leading to a chain of violent or non-violent reactions, and are circulated across borders through memes which are then used in a local context. The current complex landscape of extremism in Europe, where many extremist ideologies oppose each other, underlines the need to study this interaction in greater depth (Flemish Peace Institute, 2023). Such an in-depth understanding can provide research-based recommendations which can be informative for the adaptation of EU laws and policies.

Cumulative extremism not only intensifies threats of violent radicalisation, but also fosters social polarisation, radicalises more individuals and triggers cycles of violence. Polarised perceptions of the “in-group” as victim and the “out-group” as threat exacerbate this process. Government policies have a key role to play here: if they are to mitigate this risk, it is essential that they provide an accurate and balanced portrayal of extremism, avoiding distorted views that perpetuate confrontation (Flemish Peace Institute, 2023). Hence, the existential experiences of victims, survivors and perpetrators can inform government policies in order to address the current issues at a political level.

The diffuse nature of extremism, coupled with the use of ambiguous tactics and activity on digital platforms, makes it difficult to accurately assess its various manifestations. This, combined with the challenges faced by practitioners in identifying signs of radicalisation, can lead to both overestimation and underestimation of extremist threats. These inaccurate



assessments often reinforce tensions between opposing groups, fuel toxic narratives and generate perceptions that counter-extremism policies and measures are unevenly applied.

Media Literacy and Resilience Societies

Media Literacy and Resilience strategies emerged as core underlying themes in the overall data collection.

This approach underscores the importance of critical thinking and media literacy in combating hate speech and disinformation. Educating individuals on how to identify and challenge harmful narratives empowers communities to resist such influences effectively. In fact, the role of education in the classroom emerged as an important tool, especially in how social media can become an intelligent tool for children and young people to be able to counteract fake news.

The WP2 data, can be backed up by research that demonstrates how media literacy refers to the ability to access, analyse, evaluate, and create media in a variety of forms. This involves the development of critical thinking skills that allow individuals to understand and deconstruct the messages they encounter in the media, whether through traditional outlets or new digital platforms and digital literacy (Potter, 2016). Such skills are useful for the younger generation (Gen Z) but also for baby-boomers who might be the victims of scams and fake news that may lead to distress and anxiety.

Moreover, media literacy plays a crucial role in building resilient societies by equipping individuals with the skills needed to critically engage with media, resist manipulation, and make informed decisions. In the context of misinformation and hate speech, media literacy helps people recognize false narratives, resist divisive content, and contribute to a more informed, united society. Misinformation and disinformation can weaken societal resilience by creating confusion, undermining trust in institutions, and exacerbating social divisions. According to Wardle and Derakhsan (2017) media literacy enables people to assess the credibility of information, check sources, and verify facts, thereby reducing the spread of false information. Additionally, media literacy encourages active citizenship by giving individuals the tools they need to engage in public discourse, participate in democratic processes, and hold leaders accountable (Jenkins and Carpentier, 2013)

Media literacy can also help societies resist hate speech and extremist propaganda. By fostering critical thinking and empathy, media literacy programs encourage individuals to question harmful stereotypes and dehumanizing rhetoric (Mihailidis and Viotty, 2017). This, in turn, strengthens societal cohesion by promoting tolerance and reducing intergroup hostility.

2.1.2 Conspiracy theories in partner countries

Some authors argue that conspiracy theories are characterised by three fundamental principles: nothing happens by accident, nothing is what it seems and everything is connected. These theories presuppose the existence of a group of “conspirators”, often portrayed as evil, who act intentionally. The victims are often ordinary people who remain oblivious to the conspiracy (Ekman, 2022).



Other authors define conspiracy theories by five elements: pattern, agency, coalitions, hostility and continued secrecy. Thus, seemingly random events are always seen as intentional and organised by coalitions of actors operating in secret with bad intentions. Furthermore, conspiracy theories construct different types of conspirators, who can be both external and internal forces in society, either in power (conspiracies from above) or trying to attain power (see section below).

LGBTIQ+ community

In Europe, conspiracy theories targeting the LGBTQ+ community have become widespread, often linking these communities to harmful or bizarre claims. These conspiracies frequently portray LGBTQ+ individuals as a threat to traditional values, public safety, or morality, fuelling anti-LGBTQ+ sentiments. These conspiracy theories often play on existing societal anxieties, such as fears over child safety, religious values, and political instability, and have been spread particularly through right-wing political channels or by authoritarian regimes looking to consolidate power through scapegoating (EDMO). The persistence of these conspiracies is closely tied to underlying many factors, such as Right-Wing Authoritarianism and Social Dominance Orientation, which promote hostility toward out-groups perceived as threatening traditional norms or hierarchies. This stance emerged across the countries involved in the research.

In Poland, LGBTQ+ people have been increasingly targeted by politicians in power, religious leaders and other public figures in recent years. These attacks have manifested themselves through the dangerous and false characterisation of LGBTQ+ people as a threat to children and to Christianity (ILGA). As an example, during the first months of the COVID-19 pandemic, some members of the Catholic clergy were found to distort the reality of the virus, comparing it to a flu or linking its origin to the “sins of abortion and homosexuality”, a common idea in conspiracy theories about the LGBTQ+ community (Rachwol, 2023). In this regard, in 2019 and 2020, several municipalities in Poland self-proclaimed themselves as “LGBTQ-free zones”, an initiative backed by right-wing parties that used disinformation to portray LGBTQ+ rights as an external threat to traditional Polish values (Salvati et al, 2024).

The Polish report states that the LGBTQ+ community is one of the main targets with regard to hate speech and conspiracy theories. In an interview with a Polish expert in anti-discrimination law and LGBTQ+ rights, it was explained that some patterns are changing and new groups are being targeted by hate speech. For example, five years ago, Facebook adopted certain policies, but today the LGBTQ+ community is much more targeted than a decade ago, when they were barely talked about. The emancipation of more groups and their subsequent attack by hate-mongers underlines the need for constant updating between civil society and big tech companies.

In 2022, Italy experienced a significant increase in anti-LGBTQ+ hate speech and acts, intensified after the September elections that gave victory to the far-right Brothers of Italy party. Politicians such as Federico Mollicone and Lucio Malan made discriminatory comments against same-sex couples, while acts of violence and vandalism targeting LGBTQ+ people, such as the murder of migrant sex workers and attacks on Arcigay offices, continued. The



rejection of the Zan bill, which would have offered protection against hate crimes, has left the community vulnerable (ILGA, 2022).

Previous research in Italy (Salvati et al., 2024) has shown that people with negative attitudes towards LGBTIQ+ community tend to be older, have a lower level of education, be more religious, have a right-wing political orientation and have few personal relationships with LGBTIQ+ people. At the same time, other studies have extensively documented the presence of anti-gender movements in Italy, which use sophisticated rhetoric, based on fear and scandal, to convince the population that there is an alleged LGBTIQ+ lobby that threatens social order and morality (Salvati et al, 2024). Other studies have extensively documented the presence of anti-gender movements in Italy, which use sophisticated rhetoric, based on fear and scandal, to convince the population that there is an alleged LGBTIQ+ lobby that threatens social order and morality (Salvati et al, 2024). In this regard, the Italian anti-gender movement, which has demonstrated against the legalisation of same-sex civil unions and the implementation of gender education programmes in schools, has experienced a remarkable rise in popularity. This growth has made it a space of mobilisation and protest for politically engaged Catholic and social conservative actors. The formation and consolidation of this movement has signalled a new stage of political Catholicism, characterised both by an increase in its protest and pressure actions in the political and public spheres, and by the entry of some of its members into the electoral arena through the creation of an independent party (People of the Family) (Lavizzari & Prearo, 2018).

In the Italian report, participants called for more robust national laws, ensuring that gender-based and LGBTQ+ hate crimes are clearly included. The need for more effective enforcement mechanisms was also emphasised.

In Poland, LGBTIQ+ people have been increasingly targeted by politicians in power, religious leaders and other public figures in recent years. These attacks have manifested themselves through the dangerous and false characterisation of LGBTIQ+ people as a threat to children and to Christianity (ILGA, n.d.). As example, during the first months of the COVID-19 pandemic, some members of the Catholic clergy were found to distort the reality of the virus, comparing it to a flu or linking its origin to the “sins of abortion and homosexuality”, a common idea in conspiracy theories about the LGBTIQ+ community (Rachwol, 2023). In this regard, in 2019 and 2020, several municipalities in Poland self-proclaimed themselves as “LGBTIQ-free zones”, an initiative backed by right-wing parties that used disinformation to portray LGBTIQ+ rights as an external threat to traditional Polish values (Salvati et al, 2024).

The Polish report states that the LGBTIQ+ community is one of the main targets with regard to hate speech and conspiracy theories. In an interview with a Polish expert in anti-discrimination law and LGBTIQ+ rights, it was explained that some patterns are changing and new groups are being targeted by hate speech. For example, five years ago, Facebook adopted certain policies, but today the LGBTIQ+ community is much more targeted than a decade ago, when they were barely talked about. The emancipation of more groups and their subsequent attack by hate-mongers underlines the need for constant updating between civil society and big tech companies.



In Malta, some studies suggest that xenophobia seems to be significantly more of a problem than homophobia. One factor that could be influencing this situation is the recent legislation on LGBTIQ rights. Malta has topped ILGA-Europe's ranking for eight years for its progress on LGBTI rights and it was the only country in the C.O.N.T.A.C.T. consortium where positive attitudes towards the LGBTIQ+ community outnumbered negative ones in online comments (Assimakopoulos & Muskat, 2018). However, some negativity persists, mostly linked to religious issues. Recent incidents during EuroPride 2023, such as the vandalism of posters in Valletta and the physical assault of a queer person after the Pride parade in Gozo, show that challenges remain. In this regard, ILGA (2023) claims that there is still a gap between the law and day-to-day reality. Activists stress that while the legal framework is crucial, a change in societal attitudes is needed to ensure true inclusion and safety for LGBTI people.

In the Maltese report, a key expert pointed out that conspiracy theories tend to be cyclical and are influenced by events and political scenarios at the national level. For example, he mentioned that in September 2023, during Malta Pride, a new wave of conspiracy theories against the LGBTIQ+ community emerged, with the spread of false claims and memes about trans people, as well as insinuations that the community was trying to indoctrinate children. The expert explained that these false claims often originate in other countries, particularly in the United States, and are then repeated in Malta. As an example, he mentioned a controversy surrounding a performance by a trans person at a children's event in Valletta, in which he was accused of trying to corrupt children and performing sexual acts. Once Pride was over, the issue became less relevant, and the focus shifted to foreign communities, such as Pakistani, Filipino and Indian. According to the expert, minority groups are the main target, and conspiracy theories shift their focus depending on the news cycle. During Pride, the LGBTIQ+ community is the target, and when issues related to overbuilding arise, the focus shifts back to foreigners.

Muslim community and cultural racism

The contemporary debate around Islamophobia arose by the end of the XX century when social tensions in some European countries evidenced an increase in intolerance against Islamic communities in Europe. In the 80s, the concept was used to characterize a phenomenon of rejection and hostility that specifically affects Muslims in Europe (Bourekba, 2022). According to this author, in the 90s a series of characteristics was described to identify Islamophobic discourses such as the idea that Islam is a monolithic static bloc and an inferior (not different) religion which legitimatised anti-Muslim discourse globally. However, both the complexity of the phenomena and the diversity of contexts to which it is applied has generated many debates regarding its definition and applicability, and conspiracy theories around anti-Muslim discourses. This deliverable classifies different contents in order to differentiate between specific conceptions as religious intolerance and cultural racism.

Some authors suggest using the concept of anti-Muslim instead of Islamophobia (Halliday, Geisser, 2003; Werbner, 2005) paying attention to a specific kind of religious intolerance. In this case, Islamophobia is understood as a phenomenon that gives rise to speeches and, ultimately, to measures that validate the rejection of Islam in the public sphere in its broadest sense. Therefore, Islamophobia does not only refer to the "fear" (phobia), but also entails attitudes restrictive or prohibitive towards certain religious practices and implies certain processes of exclusion. Another relevant and growing academic school points out that not



only Muslims can suffer discrimination based on Islamophobia. Individuals whose names or appearances look like Muslim Arabs or who speak Arabic can experience discrimination or be attacked for the presence of physical appearance or ethnicity as an identity marker (Cheng, 2015; Meer and Modood, 2009 y 2012). Cultural racism refers to the kind of racism that identifies Muslims as fundamentally different and inferior for their condition of not sharing common values, culture and features presented in traditional European societies (Allen 2010; Modood 2000; Grosfoguel, 2014). According to Aman and Zempi (2020:5), Islamophobic victimisation was understood as a 'new' form of racism on the basis that there was a shift from race to religion. For this author, while the 'old' racism was based on an explicit belief in biological superiority, the 'new' racism encompasses notions of religious and cultural superiority. From this premise, specific type of discrimination that intersects anti-Muslim behaviour and racism creates boundaries between 'them' and 'us' founded upon difference rather than inferiority. As stated by Awan and Zempi (2020) cultural difference is understood as 'cultural deviance' and equated with the notion of cultural threat. This new form of racism can be interpreted as racism of 'reaction', based on the perceived 'threat' to traditional social and cultural identities and the stigmatisation of 'Other' cultures in the interests of protecting 'us' (the non-Muslim Self) from 'them'.

Understanding racism should not focus exclusively on race and thus overlook religion and culture. Thus, it is complex to determine which aspect of identity (ethnic identity and religious identity) is the object of a form of rejection or hostility (Bourekba, 2022). Consequently, this approach considers that Islamophobia is also a form of racism: it assigns Muslim people certain religious and cultural characteristics in order to denigrate, marginalize, discriminate against them and treat them as "second-class citizens" (Modood, 2018). Some reports of the OHCHR have suggested a double definition to address both phenomena. On the one hand, the link between institutional levels of Islamophobia and manifestations of such attitudes triggered by the visibility of the victim's (perceived) Muslim identity. On the other hand, Islamophobia is a 'new' form of racism, whereby Islamic religion, tradition and culture are seen as a 'threat' to Western values. Accordingly, Awan and Zempi (2020) propose a framework that indicates that victimisation can be 'ideological' and institutional (ideas and concepts that victimise individuals or groups) or it can have material consequences for those who are victimised (for example through verbal and physical abuse). This definition pays attention also to theories of cultural racism.

The application of stereotypes and a monolithic vision of Islam influenced by secular Western approaches had serious consequences for the general online content. The growth of extremist narratives against Muslim, migrants and diaspora communities is closely linked in Europe to the authorisation processes and the rise of far-right movements which are characterize by nationalist, nativist and anti-migrant discourse (Mudde, 2007).

Nativist ideology encompasses protecting the interests of native-born or established inhabitants against those that are defined like "non-natives". This idea is based on a conception of a "pure" nation-state where most parties would settle for a more attainable ethnocracy (Mudde, 2007: 138) in which "the others" are treated with hostility. The conception of otherness is central for defining identities and reflects the way individuals and groups internalise established social categories within their societies, such as their cultural (or ethnic) identities, gender identities, class identities, and so on. These social categories



shape our ideas about who we think we are, how we want to be seen by others, and the groups to which we belong. Herbert (2018) established that social identities are created through our ongoing social interaction with other people and our subsequent self-reflection about whom we think we are according to these social exchanges. Mead's work shows that identities are produced through agreement, disagreement, and negotiation with other people. We adjust our behaviour and our self-image based on our interactions and our self-reflection about these interactions. Ideas of similarity and difference are central in the far-right conception of political policy for promoting and protecting the interests of native inhabitants over those of immigrants, including the support of immigration-restriction measures; and at the same time in the self-conception of migrants as excluded from society.

The increasing visibility of Islam in Europe has also given rise to conspiracy theories that reactivate the myth of the "fifth column", which aims to alter European values, socio-demographic composition and cultures (Bourekba, 2022).

For example, the theory of Islamization maintains that the settlement of Muslim populations in Europe is part of a project for the Islamization of the continent. In the same line, the Eurabia conspiracy theory is based on the idea of the silent "Islamization" of Europe and the inclusion as a form of "positive discrimination" towards the population of Arab descent in Europe. Also, the theory of the great replacement states that the populations that migrated to Europe will end up replacing the native European populations (rhetoric invasion). Accordingly, it considers that welcoming immigration policies — particularly those impacting non-white immigrants — are part of a plot designed to undermine or "replace" the political power and culture of white people living in Western countries. For these theories, migrants seem to be part of an inverted acculturation project (Sanders, 2013). In other words, they are not integrated precisely because they do not want to, since what they seek is to impose their own way of life.

In Italy, a widely reported conspiracy theory is the so-called "Great Replacement", which argues that white Europeans are being supplanted by immigrants from non-European countries due to the actions of politicians and elites. This theory has gained traction in the context of recent waves of refugees and asylum seekers from the Middle East and Africa (Ekman, 2022). On 18 April 2023, at an event organised by the CISAL trade union in Rome, Francesco Lollobrigida, Minister of Agriculture and Food Sovereignty in the Meloni government, expressed concern about the concept of ethnic replacement, stating that Italians who have fewer children cannot be replaced by others. Moreover, before the 2022 election campaign, the Italian far right had already used hints of the great replacement theory. As an example, in a 2016 tweet, Meloni warned that while many Italians were moving abroad, immigrants were arriving in large numbers, which she considered a "rehearsal for ethnic replacement" (Tortola & Griffini, 2024).

In the Italian report, all interviewees agree that conspiracy theories fuel disinformation and recognise their dangerous potential. Some examples mentioned include conspiracies denying man's landing on the moon and anti-Semitic theories, noting that both anti-Semitism and racism in its various forms are common undercurrents in many conspiracy theories.



In Malta, Muslim migrants, as well as people from Africa or the Arab world are often the main target of discriminatory attitudes. What is striking, according to the analysis carried out by Assimakopoulos & Muskat (2018), is the tendency to confuse religion with ethnicity. For example, some Christians are classified as Muslims only because of their skin colour, while Muslims are identified as Arabs or Africans only because of their religion. This form of local xenophobia appears to be more common than homophobia, as evidenced by the higher number of discriminatory comments towards migrants in the study and the perception of respondents, who considered xenophobic comments more acceptable than homophobic comments, the latter being generally seen as completely unacceptable.

In addition to the intersection between race and gender issues, religion has also emerged as a fertile ground for hate speech and conspiracy theories as explained in the Maltese report. According to one fact-checker, about a decade ago the focus was on Islam and the Muslim community in Malta: “now we see that southern European communities are easier targets because their presence has increased. If we go back 20 years, the problem was more related to Eritreans and Somalis”. A fact-checker from the Maltese report pointed to a case of misinformation related to population growth in Malta that was widely circulated on Facebook. The claim asserted that “data from the National Statistics Office shows that Malta's population now stands at 717,000”. Such false claims reinforce ultra-nationalist and far-right emotions, fuelling conspiracy theories about the migrant population in Malta and exacerbating the gap between local residents and foreigners.

In the Spanish context, one of the most used theories affirms that if a country passes laws that are favourable to immigrants, then many immigrants will begin to arrive in that country and will be attracted by those hospitable laws. The expression call effect, illegal immigration or invasion is also used in the Spanish context linking with the replace theory. In addition to race and gender issues, religion has also emerged as a fertile ground for hate speech and conspiracy theories as explained in Malta, Italy and Poland. The hate posts pointed out Muslim/ Arab communities also involved a series of stereotypes like violence, terrorism, delinquency, homophobia, misogynist behaviour and the abuse of social assistance. In particular, the references to unaccompanied minors have been especially harmful to referring to them as criminals, drug dealers and rapists to the extent that MENA (Spanish acronym for unaccompanied minor) has become a disrespectful word. According to one fact-checker, about a decade ago the focus was on Islam and the Muslim community in Malta: “now we see that southern European communities are easier targets because their presence has increased. If we go back 20 years, the problem was more related to Eritreans and Somalis”.

Cultural racism can be also referring to discrimination against other racialised communities as ROMA community. Historically, conspiracy theories have often depicted Roma individuals as criminals or deviants, with fabricated stories about child-snatching being one such example. These negative stereotypes, combined with the community's marginalized status, have made them vulnerable to manipulation by extremist groups and political actors. For instance, during World War II, the Roma were victims of Nazi persecution, largely due to malicious propaganda that demonized them.

Historical revisionism



Far-right and radical right narratives employ historical revisionism to frame concepts of cultural, ethnic and racial nationalism. This historical revisionism is often considered non-academic or pseudo-scientific when used in a political manner, as it manipulates certain historical events to achieve objectives that lack scientific basis. Incorporated into the propagandistic frameworks of extremists, this revisionism represents a danger by reinterpreting evidence for the benefit of their political agendas. Thus, a distorted view of past events is presented. Extremists try to adapt history to their ideology and use it to further their political interests by manipulating perceptions, cognitions and emotions. Their purpose is not to shed new light on historical narratives, but to create myths that serve their ends (FUNDEA forthcoming in 2024).

Antisemitism and other ethnic identities

While anti-Semitism has been a recurrent target of right-wing extremist disinformation in Italy, the Covid-19 pandemic offered new opportunities to normalise it. Jews, along with other groups, became prime targets of hatred, while right-wing extremists used conspiracy theories to spread anti-Semitism and Islamophobia. A study by the Institute for Strategic Dialogue (2020) showed a significant increase in social media discussions about influential Jewish figures such as Bill Gates, George Soros and the Rothschilds, who were already the subject of conspiracy theories before the pandemic. Moreover, the reappearance of themes from the “Protocols of the Elders of Zion”, now rebranded in theories such as the New World Order or the already mentioned Great Replacement, reinforces the connection between conspiracy theories about Covid-19 and the far right.

The Polish report states that, apart from LGBTIQ+ people, refugees and members of ethnic and religious minorities (such as Jews) are the main targets of hate speech and conspiracy theories. Hate speech towards minorities tends to coincide, such as anti-Semitism and anti-Ukrainian sentiment. One example mentioned during one of the focused interviews was the so-called “UkroPolin” theory, which intertwines anti-Ukrainian and anti-Semitic themes. According to this conspiracy theory, Polish society is undergoing a process of Ukrainisation, where Ukrainians are coming to Poland with the intention of dismantling it and establishing a new state called Ukropolin, in collaboration with the Jewish community (Turska-Kawa & Stępień-Lampa, 2023).

2.2 Hate speech in the EU context

2.2.1 The state of implementation of the relevant EU regulations and policies

International law requires EU member states to prohibit by law any advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence. However, such prohibition must meet certain conditions to be in conformity with the right to freedom of expression and may either be done through criminalisation or appropriate civil or administrative sanctions. Similarly, the right to equality before the law and the right to non-discrimination are encoded in various instruments of international human rights law.

EU law specifically criminalises hate speech and hate motivated crime on the basis of a set of protected characteristics, through a Council of the European Union’s framework decision. EU



Member States are required to transpose it into their national criminal laws, which means that those who incur in such a conduct can be prosecuted by national authorities. In particular, the framework decision criminalises hate speech defined as “publicly inciting to violence or hatred against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin” or “publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity and war crimes” directed against members of such groups. It also prohibits hate crime defined as “any criminal offence, other than hate speech, committed with a racist and xenophobic motivation”, requiring EU Member States to take measures to ensure that racist and xenophobic motivation is considered an aggravating circumstance by their criminal laws, or such a motivation taken into consideration by the courts when setting penalties.

In addition to combating hate speech by means of criminal law, the EU has adopted legislation to combat racism, discrimination and other forms of intolerance in anti-discrimination directives; in the Victims’ Rights Directive establishing minimum standards regarding the rights, support and protection of all victims of crime; and in specific provisions on hate speech and hate crime such as the Digital Service Act, which requires digital platforms to take steps to remove illegal content to tackle disinformation and hate speech for a safe online environment.

In light of a general trend of increasing incidents of hate speech and hate related crimes registered in the EU, the European Commission has also put in place policies that complement and strengthen the rules to combat hate speech and hate crime. Such initiatives include the High-level Group on Non-Discrimination, Equality, and Diversity, set up to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age and sexual; and a voluntary Code of conduct to fight illegal hate speech online signed with major social media platforms. Other relevant actions have sought to respond to specific patterns in hate crime. For instance, responding to an escalation of discrimination and violence against Jews and Muslims in Europe, the European Commission and the High Representative of the Union for Foreign Affairs and Security Policy adopted the joint communication ‘No place for hate: A Europe united against hatred’ in December 2023, as a call to action focused against hatred on Jewish and Muslim communities.

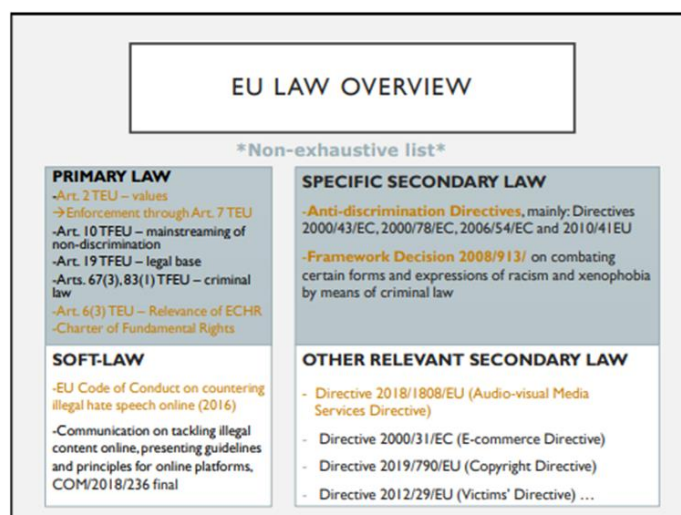
Whereas all EU Member States have criminalised hate speech based on the grounds covered by the 2008 framework decision, as well as hate crime either as a 'self-standing' offence or as an aggravating circumstance for any offence or a range of specific offences, in 2020, the European Commission began launching infringement proceedings against Member States that had incorrectly or incompletely transposed the framework decision. Member states’ criminal laws deal with hate speech and hate crime in different ways, while EU-wide rules apply only when such crimes are committed based on race, skin colour, religion, descent or national or ethnic origin. EC (2021) Communication requesting the EU Council to broaden the legal foundation for criminalising hate speech and hate crimes to encompass other categories beyond those already included in the Framework Decision and European Parliament in 2024 since the EU Council had made no progress on it.

In December 2021, the Commission proposed to extend the list of EU crimes under Article 83(1) of the Treaty on the Functioning of the European Union (TFEU) to hate speech and hate



crime. This would allow the Commission to propose a common legal framework to combat hate speech and hate crime across the EU. With this initiative, the Commission hopes to address Member States' current divergent and fragmented approaches to hate speech and hate crime and to ensure consistent protection of victims across the EU. While the Parliament has enthusiastically welcomed the Commission's proposal, Member States (in Council) have been unable to agree on the plan. EU law criminalises hate speech and hate crime, but only if it relates to a limited set of characteristics, such as race and ethnicity. The European Commission, with the European Parliament's support, would like to widen the scope of the prohibition. In December 2021, the Commission proposed to extend the list of EU crimes under Article 83(1) of the Treaty on the Functioning of the European Union (TFEU) to hate speech and hate crime. This would allow the Commission to propose a common legal framework to combat hate speech and hate crime across the EU. With this initiative, the Commission hopes to address Member States' current divergent and fragmented approaches to hate speech and hate crime and to ensure consistent protection of victims across the EU. While the Parliament has enthusiastically welcomed the Commission's proposal, Member States (in Council) have been unable to agree on the plan. At present, the EU has no competence to criminalise hate speech and hate crime based on grounds not covered by the 2008 framework decision. Article 83 TFEU could serve as a legal basis for such action should these offences become part of an exhaustive list of 'areas of particularly serious crime with a cross-border dimension resulting from the nature or impact of such offences or from a special need to combat them on a common basis' ('EU crimes').

UN Action Plan UN Strategy and Plan of Action on Hate Speech The 2019 United Nations Strategy and Plan of Action on Hate Speech defines hate speech as 'any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.' In a 2022 recommendation to Member States on combating hate speech, the Council of Europe defined hate speech as 'all types of expression that incite, promote, spread or justify violence, hatred or discrimination against a person or group of persons, or that denigrates them, by reason of their real or attributed personal characteristics or status such as "race", colour, language, religion, nationality, national or ethnic origin, age, disability, sex, gender identity and sexual orientation.'



Primary law

TEU Article 2 of the Treaty on European Union (TEU) states that '[t]he [European] Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail'. Article 10 of the Treaty on the Functioning of the European Union (TFEU) states further that 'in defining and implementing its policies and activities, the Union shall aim to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation.' Moreover, Article 19 TFEU gives the Union the power to 'take appropriate action to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation'.

Secondary law

Anti-Discrimination Directives: EU anti-discrimination directives Over the years, the EU has adopted a set of directives prohibiting discrimination on a variety of grounds. These include: Directive 2000/43/EC against discrimination on grounds of race and ethnic origin. Directive 2000/78/EC against discrimination at work on grounds of religion or belief, disability, age or sexual orientation. Directive 2006/54/EC equal treatment for men and women in matters of employment and occupation, and Directive 2004/113/EC equal treatment for men and women in the access to and supply of goods and services.

The Victims' Rights Directive was adopted in 2012 to ensure that victims of crime receive appropriate information, support and protection and are able to participate in criminal proceedings. Member States must ensure that victims are recognised and treated in a respectful, sensitive, tailored, professional and non-discriminatory manner, in all contacts with victim support or restorative justice services or a competent authority, operating within the context of criminal proceedings. Member States are obliged to carry out individual assessments to identify the specific protection needs of victims of crimes committed with a bias or discriminatory motive (Article 22 of the directive), in particular if it is related to their personal characteristics. It includes the grounds of sexual orientation, gender identity, gender expression and disability when recognising the rights of victims. Sexual and gender-based violence and hate crime are recognised as grounds for specific protection needs. In July 2023, the Commission published a proposal for a revision of the Victims' Rights Directive, including Article 22.

Directive on combating violence against women Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on combating violence against women and domestic violence lays down rules to prevent and combat violence against women. The directive requires Member States to criminalise hate speech against women alongside other forms of gender-based cyber-violence. Article 8 of the directive requires Member States to make it a crime to intentionally incite violence or hatred against a group of persons or a member of such a group, defined by reference to gender, by publicly disseminating material containing incitement online.

Specific provisions on hate speech and hate crime



Framework Decision 2008/913 (criminal law). Definition of hate speech on the basis of intentional conduct by “publicly inciting to violence or hatred directed against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin” or “publicly condoning, denying or grossly trivialising crimes of genocide, crimes against humanity and war crimes as defined in Articles 6, 7 and 8 of the Statute of the International Criminal Court...” and in “Article 6 of the Charter of the International Military Tribunal appended to the London Agreement of 8 August 1945 ...”

Directive of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law (the 2008 Council Framework Decision) criminalises hate speech and hate crime based on a range of grounds. It prohibits 'publicly inciting to violence or hatred against a group of persons or a member of such a group defined by reference to race, colour, religion, descent or national or ethnic origin'. It also requires that Member States take measures to ensure that racist and xenophobic motivation is considered an aggravating circumstance by their criminal laws, or alternatively, may be taken into consideration by the courts when setting penalties.

In 2016, the European Commission set up the High-Level Group on Hate Speech and Hate Crime. The group supports the enforcement of the 2008 Council Framework Decision on combating racism and xenophobia by means of criminal law, through guidance and sharing good practice. The work of the group is organised around four thematic areas, hate crime recording, reporting and data collection; hate crime training and capacity building for national law enforcement; hate crime victim support; and countering hate speech online. The group is made up of experts from national authorities, international organisations (Council of Europe, OSCE/ODIHR, UN), EU agencies (FRA, CEPOL) and some EU-wide civil society organisations.

The Digital Services Act (DSA) entered into force in November 2022, puts a framework of layered responsibilities in place, targeted at different types of online intermediary services. These include network infrastructure services (e.g. cloud and webhosting), online platform services (e.g. app stores and social media platforms), and services provided by very large online platforms and very large online search engines that pose particular risks in the dissemination of illegal content and societal harms. Under the DSA, all providers offering such online intermediary services in the EU have to comply with a range of obligations to ensure transparency, accountability and responsibility for their actions according to their role, size and impact in the online ecosystem. It requires all online platforms to have notice and action systems in place to allow for user reports on illegal content, including hate speech. It also requires online platforms to inform law enforcement agencies if they become aware of a serious criminal offence involving a threat to life or safety. This means that the most serious cases of online hate speech not only have to be removed, but they must also be reported to the police. Moreover, the DSA introduced strict rules for very large online platforms with more than 45 million users per month in the EU. These platforms and search engines must assess and mitigate systemic risks, also concerning the spread of illegal hate speech. Large online platforms also need to increase transparency, including by providing information on the functioning of their algorithms and recommending systems.

Soft law



EU Code of Conduct on countering illegal online hate speech (2016) is oriented to prevent and counter the spread of illegal hate speech online, in May 2016, the Commission agreed (voluntary commitment) with Facebook, Microsoft, Twitter and YouTube a “Code of conduct on countering illegal hate speech online”. In the course of 2018, Instagram, Snapchat and Dailymotion took part to the Code of Conduct, Jeuxvideo.com in January 2019, TikTok in 2020 and LinkedIn in 2021, but Twitter (now X) withdrew from it in 2023. In May and June 2022, respectively, Rakuten Viber and Twitch announced their participation to the Code of Conduct. The implementation of the Code of Conduct is evaluated through a regular monitoring exercise set up in collaboration with a network of organisations located in the different EU countries. Using a commonly agreed methodology, these organisations test how the IT companies are implementing the commitments in the Code.

To tackle the spread of illegal hate speech online, the Commission has drawn up a 'Code of conduct on countering illegal hate speech online', first introduced in 2016. Major internet companies – including Facebook, Microsoft, Twitter (but not X), 2 YouTube, Instagram, Snapchat, Dailymotion, Jeuxvideo.com, TikTok, LinkedIn, Rakuten Viber and Twitch – have signed up to the Code of Conduct. By signing up, these companies commit to ensuring that online platforms do not offer opportunities for illegal online hate speech to spread virally. More specifically, they agree to put in place clear and effective processes to review notifications regarding illegal hate speech on their services so they can remove or disable access to such content. Upon receipt of a valid removal notification, companies commit to review such requests against their rules and community guidelines and, where necessary, national laws transposing Framework Decision 2008/913/JHA, with dedicated teams reviewing requests. Moreover, they agree to review the majority of valid notifications for removal of illegal hate speech in less than 24 hours and remove or disable access to such content, if necessary. The European Commission regularly monitors and evaluates the implementation of the Code of Conduct, in collaboration with a network of organisations located in the different EU countries. Using a commonly agreed methodology, these organisations test how IT companies are implementing the commitments in the Code. In November 2022, the Commission deplored a 'slow-down in progress', EU High Level Group on Hate Speech and Hate Crime. In 2016, the European Commission set up the High-Level Group on Hate Speech and Hate Crime. The group supports the enforcement of the 2008 Council Framework Decision on combating racism and xenophobia by means of criminal law, through guidance and sharing good practice. The group organises in-depth expert discussions on horizontal issues common to the fight against hate speech and hate crime and has developed practical guidance, standards and tools and seeks to build capacity to improve responses at the national and local level. The work of the group is organised around four thematic areas, hate crime recording, reporting and data collection; hate crime training and capacity building for national law enforcement; hate crime victim support; and countering hate speech online. The group is made up of experts from national authorities, international organisations (Council of Europe, OSCE/ODIHR, UN), EU agencies (FRA, CEPOL) and some EU-wide civil society organisations. However, a meeting of the High-Level Group on Combating Hate Speech and Hate Crimes held in October 2023 concluded that the Code of Conduct had 'become an efficient and successful instrument to tackle online hate speech over the years'. To take account of the provisions of the Digital Services Act (see above), the Commission has started to discuss the features of a forthcoming enhanced code, the 'Code of Conduct+', together with the signatories of the original Code of Conduct, national authorities and civil



society organisations. Under the new Code of Conduct+, the Commission wants online platforms to enhance prevention and anticipate threats. For this reason, the Commission proposes to establish exchange fora and a knowledge hub, to enable platforms to receive timely information on possible trends and developments in hate speech or threats to freedom of expression.

The EU Anti-racism Action Plan (2020-2025) establish a committee of experts with preparing a draft recommendation on combating hate crime (27), and in November 2023, adopted a study that provided guidance to Member States and other actors on preventing and combating hate speech in times of crisis (28).

Limits to freedom of expression:

Article 20 (2) of the International Covenant on Civil and Political Rights (ICCPR) obliges States parties to prohibit by law “any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence”, while its article 19 protects freedom of opinion and expression. At a global level, guidance from the Office of the High Commissioner for Human Rights (OHCHR) has sought to strike a balance that responds to the need of addressing the genuine harm of hate speech and the need to prevent its instrumentalization to target peaceful critics, particularly by authoritarian governments.

2.2.2 General overview in partner countries

This section seeks to understand, through participants' perceptions, how legislation in partner countries compares with other EU member states, especially in relation to the protection of vulnerable groups.

In the Italian context, respondents express a clear dissatisfaction with the current legislation, considering it insufficient to protect vulnerable groups, such as gender-based hate crimes. A critical point in this assessment is the failure of the DDL Zan law, which was intended to provide a stronger legal framework to combat violence and discrimination based on gender identity and sexual orientation. The lack of a viable alternative to this legislation is of concern to participants, who feel that the absence of specific protections allows acts of hatred in various forms to be perpetuated. In addition, it is noted that Italian legislation does not adequately address hate speech in the digital sphere. Compared to countries such as Germany and France, where laws are stricter and effectively enforced, Italy is seen as ‘lagging behind’ in terms of regulation and enforcement. Participants highlight a significant disparity in legal frameworks across Europe, underlining the urgency of implementing reforms to strengthen the legal tools available to combat hate speech and disinformation.

The assessment of Polish legislation also reveals notable shortcomings. Participants agree that existing laws do not provide adequate support in the fight against hate speech. One area of concern is that the Polish Penal Code does not include explicit protections for LGBTIQ+ people, who are particularly vulnerable to various forms of violence and discrimination. Despite several attempts to amend the penal code to include such protections, no significant progress has been made so far. Jurisprudence in Poland also faces criticism, as it is considered inconsistent and often inadequate in its response to hate crimes. A shocking testimony from a victim illustrates the lack of understanding in law enforcement, where acts of direct threat



are downplayed or misinterpreted by judicial authorities. This suggests that not only is there a need to improve the legislation itself, but also to train legal professionals to better understand the nature of hate speech in the digital context and its impact on victims.

An additional challenge is the preparedness of law enforcement agencies to handle hate crimes, especially in the online environment. Many officers are reportedly not technologically equipped to accept and process reports of hate crimes committed via the internet. For example, some police are unable to open links to materials that could be evidence in a case, indicating a lack of adequate infrastructure to deal with hate crimes in the digital age. The need for training and resources is urgent, as it is recognised that a better understanding of the language used in hate speech is critical for effective prosecution.

In Malta, although the context is different, a normalisation of hate speech is also observed, with a particular focus on migrant communities. The perception that hate speech is present on social media on a daily basis resonates with experiences in Italy and Poland. One participant noted that this phenomenon has been normalised both by the public and by the victims themselves, who often feel that they have more pressing problems. Furthermore, it was reported that the digital platforms in charge of handling abuse are often located in countries such as Vietnam and the Philippines, which limits their understanding of Malta's cultural specificities.

In both cases, Italy and Poland agree on the need to strengthen national laws. In Italy, a more robust approach to online hate speech is advocated, with a particular emphasis on ensuring that hate crimes based on gender and sexual orientation are explicitly covered by law. In Poland, measures such as the creation of specialised hate crime teams within the police and prosecutor's office are suggested, as well as the need to update the General Prosecutor's Office guidelines to address changes in the digital environment since the last revision in 2014.

While there is an interrelationship between hate speech and disinformation, participants underline that disinformative content is often more difficult to remove from the internet due to its not always illegal nature. This poses an additional challenge in the fight against hate speech, as disinformation can fuel hate narratives and contribute to an environment of hostility towards marginalised groups. In this regard, both Italy and Poland need not only a more robust legal framework, but also a concerted commitment to address these problems at multiple levels, from legislation to education and public awareness.

In Italy, migration was a focus of public debate during their respective election campaigns. Right-wing parties, politicians and public officials, as well as explicitly xenophobic movements, fuelled hostility towards minorities and migrants and refugees in order to gain political support. Prohibition of hate speech in criminal law: “instigating” or “provoking violence” based on protected grounds or “public instigation to disobedience of public order or hatred amongst social classes” (incitement); “race, ethnic origin, nationality or religion” (protected characteristics, hate speech is recognised on these grounds). Also, there are criminal offences that might indirectly criminalise some types of ‘hate speech’, for example Defamation is a criminal offence; and the courts have applied the crime of “criminal conspiracy” to a case concerning the promotion of neo-Nazi ideology online, and the coordination of an online group, also include offences criminalising the denial of historical events or the endorsement



of certain ideologies. These include, for example, “apology of fascism” and “denial of the Shoah or of genocide crimes, war crimes and crimes against humanity”. https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf from discrimination. Prohibition of hate speech in criminal law: “the use of unlawful threats directed against groups or individuals” or “public insult of group of people or individuals” on protected grounds (incitement). Also, there are criminal offences that might indirectly criminalise some types of ‘hate speech’: Defamation is a criminal offence. Poland (“insult of religious beliefs or offending religious feelings”. https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf

In Malta, although the context is different, a normalisation of hate speech is also observed, with a particular focus on migrant communities. The perception that hate speech is present on social media on a daily basis resonates with experiences in Italy and Poland. One participant noted that this phenomenon has been normalised both by the public and by the victims themselves, who often feel that they have more pressing problems. Furthermore, it was reported that the digital platforms in charge of handling abuse are often located in countries such as Vietnam and the Philippines, which limits their understanding of Malta's cultural specificities.

In both cases, Italy and Poland agree on the need to strengthen national laws. In Italy, a more robust approach to online hate speech is advocated, with a particular emphasis on ensuring that hate crimes based on gender and sexual orientation are explicitly covered by law. In Poland, measures such as the creation of specialised hate crime teams within the police and prosecutor's office are suggested, as well as the need to update the General Prosecutor's Office guidelines to address changes in the digital environment since the last revision in 2014.

Finally, while there is an interrelationship between hate speech and disinformation, most participants in the research underline that disinformative content is often more difficult to remove from the internet due to its ‘not always’ illegal nature. This poses an additional challenge in the fight against hate speech, as disinformation can fuel hate narratives and contribute to an environment of hostility towards marginalised groups. In this regard, both Italy and Poland need not only a more robust legal framework, but also a concerted commitment to address these problems at multiple levels, from legislation to education and public awareness.

The Spanish legal framework was taken as a starting point for discussing the existing mechanism for fighting against hate crime. According to the participants’ knowledge, Spanish law identifies hate crimes as criminal acts that are committed based on prejudice.

Hate crime is founded on two basic elements. On the one hand, the act has to constitute a criminal offense. On the other hand, the perpetrator’s prejudice towards the victim has to be related to belonging to a vulnerable group. In other words, the hate crime legal definition assumes that the victim (or target) is intentionally chosen by the perpetrator because of their (real or presumed) membership in a group considered especially protected against hate crimes. Therefore, the distinctive elements of these crimes are:



- (1) The victim (or victims) has a symbolic condition because they are attacked for their characteristics or the characteristics they seem to have;
- (2) The intention of this type of violence is not only to attack the victim but also to transmit a message of rejection, hostility, and intimidation to the entire group to which the victim belongs.

As one of the FG experts noticed, the legal definition of hate crime in Spanish law is in line with the main OSCE Hate Crime Reporting (OSCE, 2021) definition, which described hate crimes as criminal acts motivated by bias or prejudice towards particular groups of people. The OSCE definition of hate crimes also comprises these two main elements: a criminal offense and a bias motivation. Thus, a hate crime has taken place when a perpetrator has intentionally targeted an individual or property because of one or more identity traits or expressed hostility towards these identity traits during the crime.

However, the Spanish Criminal Code contains a lack of clear and well-defined regulatory legislation on this matter. Two legislative models are represented in the Criminal Code that differentiates between a reaffirmation of the principle of equality; and a preventive discriminatory model for reinforcing the protection of vulnerable groups against discrimination. As a result, hate crimes are considered in the Spanish legal system from two perspectives. First of all, crimes whose criminal responsibility is aggravated by the author's hatred or criminal prejudice towards a certain personal condition of his or her victim following the discrimination principle. In contrast, those crimes whose criminal responsibility is aggravated because it produces an intimidating effect on the group to which the victim belongs due to their specific personal conditions and regardless of the perpetrator's motive was (commonly called 'strictly hate crimes'). In this vein, the Criminal Law (article 510) penalizes hate speech itself or the emission of any message or content against human rights promoting violence against members of a group in its version of publishing speeches of that nature, broadcasting them on the Internet, favouring the publication, edition, diffusion, etc. In other words, article 510 condemned aggressions against a person whose main purpose is to intimidate the members of a group materialized as hate speech. In addition, the Criminal Law refers to discriminatory access to public services (511), business sphere (512), illicit association to commit discriminatory acts (515.4), and crimes against freedom of conscience and religious feelings (522-525).

Nevertheless, other references in the same law can be applied to discriminatory acts. Indeed article 22.4 of the Spanish Criminal Law is considered an instrument for aggravating circumstances of discrimination or hate motives. Also, article 22.4 regulates an aggravating circumstance for crime against moral integrity by having been committed with a motive of hatred or discrimination (article 173). Finally, the Criminal Law contains disseminate articles about crimes of threats against an ethnic, cultural, or religious group, a social or professional group, or any other group of people (170.1); crimes against discrimination in the workplace (314); and crimes of genocide and crimes against humanity (607 and 607 bis).

Related to legal conceptualizations of hate, some participants express their concern over the tendency to identify 'hate crimes' with 'hate speech'. As we have described in the previous point, the incitement to hatred in the Spanish law (hate speech) is part of that list of crimes included under the umbrella of what is conceptually called "hate crimes". This means that



article 510 about hate crimes is also applied to crimes of incitement to hatred, violence, or discrimination. As one of the participants (LEAS) mentioned, the criteria for applying the law depends on the motivation of the crime so the most important element for distinguishing hate crimes or hate speech from other kinds of criminal acts is to prove motivation.

Article 510 of the Criminal Law also compiles a closed list of the biased motivations for committing hate crimes in Spain. This categorization is based on the human rights normative body subscribed by Spain. Therefore, the law constraints to the following list of reasons that individuals can allocate as hate crimes: racist acts, anti-Semitic reasons, ideology, religion or beliefs, family situation, belonging to an ethnic group, race or nation, national origin, sex, sexual orientation or identity and reasons of gender, illness or disability. Protective and promotive factors and community resilience techniques against the drivers of hate speech and disinformation, as well as methods/tools to reduce (multiple) victimisation(s), ensure victims' protection and promote sensitiveness among victims.

In Italy, immigration was a focus of public debate during their respective election campaigns. Right-wing parties, politicians and public officials, as well as explicitly xenophobic movements, fuelled hostility towards minorities and migrants and refugees in order to gain political support. Prohibition of hate speech in criminal law: "instigating" or "provoking violence" based on protected grounds or "public instigation to disobedience of public order or hatred amongst social classes" (incitement); "race, ethnic origin, nationality or religion" (protected characteristics, hate speech is recognised on these grounds). Also, there are criminal offences that might indirectly criminalise some types of 'hate speech', for example Defamation is a criminal offence; and the courts have applied the crime of "criminal conspiracy" to a case concerning the promotion of neo-Nazi ideology online, and the coordination of an online group, also include offences criminalising the denial of historical events or the endorsement of certain ideologies. These include, for example, "apology of fascism" and "denial of the Shoah or of genocide crimes, war crimes and crimes against humanity".

A recent study by FRA examined popular social media platforms, trying to identify cases of hate speech in four selected countries, including Italy, indicated that women was the group facing the most hateful speech with an incidence double higher than the other groups considered for the study (People of African Descent, Jews, Roma) in pp. 57-58 https://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-online-content-moderation_en.pdf. In Italy, an administrative bill is currently being discussed that would create penalties for websites that do not remove hate content from their web pages (Italy, Administrative proposal – measures to prevent and combat the spread of hate manifestations through the internet (Proposta di Legge – Misure per la prevenzione e il contrasto della diffusione di manifestazioni d'odio mediante la rete internet), Camera dei Deputati, No 2936, 10 March 2021.)

In both Poland, the national government has been very vocal in their anti-EU and anti-migrant rhetoric. additionally have a long-standing problem with ensuring racial, ethnic and religious minorities, and LGBT people are protected from discrimination. Prohibition of hate speech in criminal law: "the use of unlawful threats directed against groups or individuals" or "public insult of group of people or individuals" on protected grounds (incitement). Also, there are criminal offences that might indirectly criminalise some types of 'hate speech': Defamation is



a criminal offence. Poland (“insult of religious beliefs or offending religious feelings”. https://www.article19.org/wp-content/uploads/2018/03/ECA-hate-speech-compilation-report_March-2018.pdf

Poland is one of the countries against which the European Commission (EC) launched infringement proceedings for incorrectly or incompletely transposing the rules on combating racism and xenophobia by means of criminal law (Framework Decision 2008/913/JHA). In a letter of formal notice sent in 2021, the EC had determined that the Polish criminal legal framework failed to correctly transpose hate speech inciting to racist and xenophobic violence and restricts the scope of the criminalisation of incitement to hatred. Additionally, it found that Poland had incorrectly transposed the criminalisation of specific forms of hate speech, by omitting the conduct of gross trivialisation of international crimes and the Holocaust and by restricting the criminalisation of the denial and condoning of those crimes only to cases where such crimes were committed against Polish citizens. An additional letter was sent to Poland in 2023 finding transposition issues in the new information provided by Poland, in addition to concerns already raised.

3 QUALITATIVE and QUANTITATIVE ANALYSIS OF EXISTING BIG DATA

REC Project Hatemeter

The European project “Hatemeter - Hate speech tool for monitoring, analysing and tackling Anti-Muslim hatred online” (hereinafter also referred to as “Hatemeter”, project reference: 764583) aimed at systematising, augmenting and sharing knowledge on Anti-Muslim hatred online, and at increasing the efficiency and effectiveness of non-governmental organisations (NGOs) in preventing and tackling Islamophobia at the EU level.

Within the project, a database of social media posts in English, Italian and French was collected to analyse the problem of Islamophobia and online discourse on the topic. Specifically, tweets and YouTube comments containing offensive keywords such as “banislam” were retrieved between September 2019 and mid 2020. Overall, the HATEMETER database contained around 3.1 million tweets (including replies and retweets) and 170,000 YouTube comments with Islamophobic content, with a prevalence of English data. The database has been currently deleted for privacy reasons, but nevertheless it was only accessible for navigation to the consortium partners through the (password-protected) HATEMETER platform.

Covid19 Infodemics Observatory

The COVID-19 Infodemics Observatory was established to analyse the spread of information during the pandemic using large-scale social media data. This project focused primarily on Twitter as a data source, collecting and analysing tweets related to COVID-19 from January 2020 onwards.

The research team identified a set of key hashtags and keywords associated with the pandemic, including “coronavirus,” “covid19,” and “sarscov2,” among others. Using Twitter's Filter API, they collected tweets in real-time, aggregating results daily. This approach allowed



for continuous monitoring of COVID-19 related conversations across multiple languages, starting from January 22, 2020, when the international community began paying increased attention to the outbreak.

The data collection process was extensive, gathering over a billion tweets between January and March 2023. However, due to Twitter's volume limitations, the team could collect a maximum of about 4.5 million messages per day when global attention to COVID-19 intensified.

To analyse the geographical distribution of tweets, the researchers used the user's self-declared location field, geocoding it with the ArcGIS API. This process allowed them to assign latitude and longitude coordinates to approximately 56% of users, although they had to filter out a significant number of incorrectly attributed locations.

A key aspect of the observatory's work was the classification of information sources. The team developed a comprehensive database of web domains, categorizing them based on their reliability and potential for harm. They assigned a "Harm Score" to different categories of sources, ranging from scientific content (lowest harm score) to conspiracy theories and junk science (highest harm score). This classification system allowed for a nuanced analysis of the types of information spreading during the pandemic.

The observatory's approach, while comprehensive, faced several limitations. These included biases inherent to Twitter's user demographics, limitations in capturing non-Latin character languages, and the static nature of their keyword list for data collection. Additionally, the reliance on Western-centric fact-checking sites potentially skewed the analysis towards English-language content.

AI4TRUST

The AI4TRUST project has undertaken a significant effort to analyse existing big data collected from social media platforms, focusing primarily on YouTube and Telegram. This analysis aims to monitor information flows and detect potential disinformation across these platforms.

For YouTube, the project implemented a sophisticated data collection pipeline allowing for efficient gathering and storage of video metadata, comments, and associated information through the YouTube Data API. The collection process respects privacy concerns and adheres to YouTube's research program requirements, ensuring ethical data usage.

Telegram API does not allow for keyword-based research of content. For this reason, the Telegram data collection employed a unique snowballing technique to discover relevant public channels. Starting from an initial seed of channels identified through keyword searches, the team expanded their dataset by leveraging Telegram's channel recommendation feature and analysing message forwarding patterns. This approach has led in the exploratory phase to the discovery of over 100,000 public channels, with approximately 86 million messages collected from 3,500 of these channels.

Both data collection efforts prioritize privacy and ethical considerations. The team implemented robust pseudonymization and anonymization techniques for user data, focusing on publicly available information while avoiding the collection of personal or copyrighted



content. This approach aligns with the platforms' terms of service and broader data protection regulations.

The project's technical infrastructure is built on a combination of cutting-edge technologies, including Apache Kafka for real-time data streaming, Minio for data lake storage, and Apache Iceberg for efficient data formatting. This infrastructure allows for seamless processing and analysis of the large volumes of data collected.

4 THE STRATEGIES AND GOOD PRACTICES TO FACE HATE SPEECH AND DISINFORMATION

Multiagency cooperation

Tackling hate crimes involves a vast number of procedures including reporting, investigation, prosecution, and prevention. Therefore, coordination between CSO, law enforcement, and judicial actors taking part in this process is essential for countering hate crime. In this regard, participants noticed a lack of multi-agency cooperation as well as an absence of common spaces in which public authorities and civil society can assign responsibilities and competencies and coordinate activities to establish a common definition of hate crimes, harmonized procedures, and fluid dialogue for building trustful relationships.

As a matter of example, participants established a debate about the amount of underreporting of hate crimes. In the case of Spain, 89,2% of respondents to the FRA Second European Union Minorities and Discrimination Survey 2017 that suffered hate crimes did not report the incident (FRA, 2017). One of the points that representatives from all the countries of the vulnerable communities mentioned were the absence of trust of victims in public authorities. Particularly, LGTBIQ+ communities remarked that public bodies in charge of reporting, investigating, and prosecuting hate crimes are not trained in LGTBIQ+ issues. Thus, victims are mainly supported by associations of their communities due to a lack of awareness of public services.

However, it should be mentioned that Spain implemented some measures to improve public bodies' capacities. First, the Spanish Prosecutor's Office has created specialized units in hate crime and discrimination whose main objective is coordinating actions of the prosecutor's offices and promoting training and awareness for an effective response to these criminal acts. Second, procedures for reporting, investigating, and prosecuting hate crimes were unified in 2022. The Investigation Unit of the National Police is in charge of unifying procedures through active involvement in any case of hate crime that is reported in any city in Spain. In addition, it is necessary to improve community relations and interactions that lead to more successful investigations, not only in cases of hate crimes but also in other matters in which police require community assistance. Finally, it is key to enhancing intelligence such as instrument of information and cooperating between public bodies and CSOs.

Victim-centred approach



The victim-centered approach in addressing hate crimes is crucial for various reasons, significantly enhancing the response to such incidents and promoting justice for affected individuals and communities. A victim-centered approach focuses on the needs and perspectives of the victims, empowering them to participate actively in the process. This approach helps victims regain a sense of control and agency, which is often stripped away during a hate crime. It emphasizes the importance of listening to victims, validating their experiences, and ensuring they feel supported throughout the legal and recovery processes.

Victims of hate crimes often hesitate to report incidents due to fear of discrimination, disbelief, or lack of support from law enforcement. By prioritizing the victim's experience, agencies can build trust within marginalized communities, encouraging more individuals to come forward. This increased reporting can lead to better data collection on hate crimes, which is essential for understanding the scale of the problem and implementing effective measures.

This approach also involves providing tailored support services, including counselling, legal assistance, and community resources. Such services can address the specific needs of victims, helping them cope with trauma and navigate the aftermath of hate crimes. For example, organizations may offer specialized resources for different marginalized groups, acknowledging the unique challenges they face.

This approach requires to train public bodies and other relevant actors in cultural competence and sensitivity towards the diverse backgrounds of victims. Understanding the cultural context can significantly impact how victims perceive support services and their willingness to engage with authorities. This fosters a more inclusive environment where all victims feel valued and respected.

By centering the experiences of victims, a comprehensive understanding of hate crimes emerges, recognizing not just the legal implications but also the social, psychological, and cultural dimensions. This holistic view can inform better policy-making and community initiatives aimed at preventing hate crimes and promoting social cohesion.

Promote training and awareness

Promoting awareness and training regarding hate speech and hate crimes is crucial for fostering inclusive communities and preventing violence. Here are several effective practices to achieve this:

- Educational Workshops and Training Programs
- -Community Engagement tailoring different audiences
- -Online Awareness Campaigns
- Collaboration with Local Organizations and improving partnerships working with local NGOs, advocacy groups, and community organizations

Particularly, training programs provide participants with knowledge about the impact of hate speech and hate crimes, including the legal definitions and societal consequences. Education fosters a deeper understanding of the issues, helping individuals recognize hate speech when they encounter it and understand its harmful effects on individuals and communities.



Awareness initiatives often incorporate stories from victims of hate crimes, which can humanize the statistics and foster empathy. By understanding the lived experiences of those affected by hate, community members are more likely to stand in solidarity with marginalized groups, thus creating a supportive environment for all. Awareness campaigns that promote diversity and inclusion help to combat stereotypes and misinformation that often underpin hate speech. By fostering dialogue and understanding among different cultural, racial, and religious groups, communities can work together to create a more inclusive society, making hate speech less socially acceptable.

5 CONCLUSION

The HATEDEMICS project tackles the urgent and complex issue of the spread of hate speech and disinformation, particularly within the context of digital environments. Through this desk research, supported by national interviews and focus groups, we have explored the underlying social, psychological, and legal frameworks that explain the proliferation of these harmful phenomena. The research emphasizes the role of social identity, framing, and media dynamics in fuelling hate speech and disinformation, while also pointing to the psychological and societal impacts these narratives have on individuals and communities, particularly in vulnerable or marginalized groups.

The national focus groups and interviews conducted across different EU member states have provided critical insights into the country-specific challenges related to hate speech and disinformation. These national findings illustrate the context-specific nature of hate speech and disinformation across the EU, emphasizing the need for tailored approaches that address the unique social, political, and cultural dynamics in each country. They also underscore the importance of fostering media literacy and digital resilience in local communities to equip individuals with the skills necessary to critically assess and challenge false narratives.

The legal and regulatory landscape remains a critical area for improvement. While the European Union has implemented directives and frameworks to combat hate speech and disinformation, national implementation varies, leading to gaps in protection and enforcement. Harmonizing these frameworks across the EU is essential to ensure consistent protection for all citizens, regardless of the member state they reside in.

Looking ahead, the development of AI-based solutions will play a pivotal role in detecting and mitigating the spread of hate speech and disinformation online. However, these technologies must be designed and implemented in ways that respect fundamental rights and freedoms. Close collaboration between policymakers, technology developers, and civil society will be crucial to ensure that these tools are both effective and ethical.

In conclusion, this report provides a comprehensive theoretical and empirical foundation for the next stages of the HATEDEMICS project. The findings from national focus groups and interviews offer valuable insights into the real-world experiences of individuals affected by hate speech and disinformation, complementing the broader theoretical analysis. As the project moves forward, these insights will inform the development of practical, AI-driven solutions and educational initiatives aimed at combating hate speech, fostering inclusion, and



promoting tolerance across Europe. Ultimately, the success of HATEDEMICS will depend on its ability to integrate research, technology, and policy to create a more cohesive, informed, and resilient European society.

6 RECOMMENDATIONS

From the above, it becomes clear that it is essential to improve responses to hate speech and hate crimes to foster safe and inclusive communities. Here are several recommendations based on best practices and expert insights:

Enhanced training in cultural competency and victim-centered approaches

1. Training programs should focus on cultural awareness and the specific challenges faced by marginalized communities. This training can help law enforcement officials understand the context of hate crimes better and respond more effectively. Regular workshops and refresher courses can ensure ongoing education.
2. Incorporating victim support and trauma-informed practices into law enforcement training can improve interactions with victims, making them feel heard and supported.

Develop community engagement initiatives

1. Partnerships with local organizations: Collaborating with community organizations can help law enforcement build trust within marginalized groups. These partnerships can also facilitate the sharing of resources and information regarding hate incidents and the support available for victims.
2. Public Awareness Campaigns: Implementing community-wide campaigns that educate the public about hate speech, its impacts, and how to report incidents can increase awareness and encourage proactive responses.
3. Support and Resources for Victims
 - Accessible Reporting Mechanisms: Establishing clear, accessible channels for reporting hate crimes and incidents is vital. These should include options for anonymity and the ability to report online.
 - Holistic Support Services: Providing comprehensive support services, such as counselling, legal assistance, and social services for victims, can help them recover and rebuild after experiencing hate incidents.
4. Enhance multi-agency cooperation

Legal Framework and Policy Reform

1. Stronger legislation: Advocating for the introduction or strengthening of hate crime legislation can deter future incidents. Laws should clearly define hate crimes and outline appropriate penalties, ensuring that perpetrators are held accountable.



2. Data collection and transparency: Improving the collection of data related to hate crimes can help authorities understand the scope of the issue and inform policy decisions. Transparent reporting can also reassure communities that their concerns are being taken seriously and improve victims' protection

Promoting bystander intervention

1. Education on Safe Interventions: Training community members on how to safely intervene when witnessing hate speech or hate crimes can empower them to act. This can include strategies for de-escalation or providing support to victims.
2. Building Solidarity Networks: Encouraging solidarity among different community groups can help create a united front against hate. Initiatives that promote intergroup dialogue and collaboration can strengthen community ties.



REFERENCES

- Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggression, sensitivity to aggression, and aggression-related expectations in children: A longitudinal study. *Developmental Psychology*, 37(5), 771-781.
- Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *Journal of Personality and Social Psychology*, 59(5), 829-841.
- Crenshaw, K. (1995). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241-1299.
- Delgado, R., & Stefancic, J. (2017). *Critical race theory: An introduction* (3rd ed.). New York University Press.
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51-58.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harper & Row.
- Iyengar, S. (1991). Is anyone responsible? How television frames political issues. *The University of Chicago Press*. McIntyre, L. (2018). *Post-truth*. MIT Press.
- Noelle-Neumann, E. (1974). Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, 24(2), 43-51.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103-122.
- Suiter, J. (2016). Post-truth politics: The rise of the alternative facts. *International Journal of Public Opinion Research*, 28(4), 527-535.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Brooks/Cole.
- Wardle, C., & Derakhsan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
- Zillman, D. (1983). Transfer of excitation in emotional behavior. *Psychological Bulletin*, 94(3), 438-463.
- Zillman, D. (1994). Enhancing the enjoyment of suspenseful films through misattribution. *Journal of Communication*, 44(4), 164-179.

