A framework on media-educational initiatives to contrast online hate speech[1]

Un framework sulle iniziative media-educative per contrastare i discorsi d'odio online

di
Barbara Bruschi
barbara.bruschi@unito.it
Manuela Repetto
manuela.repetto@unito.it
Melania Talarico
melania.talarico@unito.it
Università di Torino

Abstract:
Online hate speech is receiving more and more attention in the present society as a great concern: this phenomenon is rapidly increasing, mainly through the widespread use of social media that expose young people to dangerous content. This makes it urgent to define an integrated and interdisciplinary strategy for preventing and contrasting hate speech since childhood. Starting from an analysis of international literature on initiatives and approaches undertaken to educate youth to face online hate speech, this study aims at defining a theoretical framework that could inspire media education initiatives providing students with a sense of civic responsibility and actively involving them. The

---

[1] B. Bruschi, M. Repetto and M. Talarico conceived of the presented idea. B.B. developed paragraph 1. and 5. M.R. developed paragraph 4. M.T. developed paragraphs 2 and 3. All authors discussed the text and revised the final manuscript.

framework was designed adopting the perspective of socio-cultural ecosystems; it maps the possible strategies and actions to prevent and contrast hate speech.

Keywords: Online hate speech; media and information literacy; formal and informal education; active citizenship.

Abstract:
I discorsi d'odio online stanno attualmente destando una preoccupazione crescente: questo fenomeno sta infatti crescendo rapidamente, principalmente attraverso l'uso massiccio dei social media che espongono le giovani generazioni a contenuti pericolosi. Diviene dunque urgente definire una strategia integrata ed interdisciplinare per prevenire e contrastare i discorsi d'odio online fin dall'infanzia. Partendo da un'analisi della letteratura internazionale su iniziative ed approcci impiegati per educare i giovani a fronteggiare i discorsi d'odio online, questo studio mira a definire un framework teorico che possa ispirare iniziative media-educative atte a promuovere negli studenti il senso di responsabilità civile e a coinvolgerli in modo attivo. Il framework, che è stato sviluppato secondo la prospettiva degli ecosistemi socio-culturali, evidenzia le possibili strategie ed azioni per prevenire e contrastare l'odio online.

Parole chiave: Discorsi d'odio online; media and information literacy; formazione formale ed informale; cittadinanza attiva

1. Introduction

Digital transition presents great opportunities for socio-economic improvements but, at the same time, it imposes important challenges such as new skill requirements (Morandini et al., 2020). Among the broad set of new life skills that educational systems should be able to promote, critical media literacy can be considered essential to counter emerging phenomenons of online discrimination among which the most known are online hate speech and cyberbullying. Education plays an essential role not only to contrast this phenomenons, but mainly to prevent their diffusion, fostering the development of critical thinking and responsibilty. According to the United Nations Strategy and Plan of Action on Hate Speech, education has a transformative power to promote the values of peace, inclusion and social justice for all, in alignment with the UN's 2030 Agenda. In this context, education on critical media literacy becomes an urgent need of the present time, in order to make anybody able to observe, analyse, evaluate, create and fully – but safely - participate to the digital life. The whole population should become acutely aware of what writing and making public a comment means and of its possible consequences (Bagnato, 2020). In this way, taking part to media educational initiatives children, young people and citizens can become not only digital literate with basic skills, but also and above all critically digital literate.

At European level, the DigComp Framework for Citizens moves in this direction, identifying the key areas of digital competence and providing guidance for citizens to confidently, critically and responsibly use everyday digital technologies. In the Digcomp Framework media literacy isn't limited to the mere use of technology and tools, but involves the necessity for citizens to develop critical

thinking skills to exercise judgement, to distinguish real information from opinion and to be capable of analysing complex realities (Vuorikari et al., 2022).

In this contribution, after a brief literature background to frame the concept of online hate speech and its implications, a framework based on the theory of socio-cultural ecosystems will be defined to inspire media educational initiatives aimed at prevent online hate speech. Since the role of social media is pervasive nowadays as a constitutive environment in which our identities are formed and where a large part of the communication takes place (O'Neill, 2015), media ecology is the horizon in which educational studies on how to prevent online hate speech could be framed. This framework maps the possible strategies and actions to prevent and contrast online hate speech considering the involved actors and social structures as well as the levels of reciprocal influence. The taxonomy deriving from this approach considers as main coordinates the axis digital-real life and the formal-informal continuum, taking into account that agency developed within a sphere affects the other side and that the main focus of media-educational initiatives should take into account how to bridge these interconnected spheres.

2. Background on online hate speech

The phenomenon of online hate speech has proliferated through digital communication in the last years and exarcebated first during the COVID-19 pandemic and presently with the global crisis resulting from war in Ukraine. The research on prevention and contrast of online hate speech has to face with a very complex field, which is full of dichotomies and contradictions starting from the concept itself of hate speech. A first contradiction lies in the definition of the term "online hate speech" itself and can be found in the blurred boundaries between freedom of speech and limitation of speech. In the United States the Constitution, through its First Amendment, protects hate speech in social media whereas it doesn't intend threaten or deliberately harm anybody; however, the various attempts to define what constitutes a threat and at what extent to curtail hate speech represents a violation appear too ambiguous or inconsistent. There is the need to draw a clear demarcation between protected speech and cyberbullying (Udoh-Oshin, 2017), also to make people more aware and to educate them to prevent and contrast this phenomenon. Another antinomy is the coexistence of two type of messages that are conveyed by hate speech (Waldron, 2012). The former is for the targeted group or minority and tries to dehumanise them; the latter is directed at people with similar views, to build and reinforce a sense of a group and to distinguish it from the others, perceived as different and as people to discriminate. Thus hate speech paradoxically both divides and unites at the same time (Gagliardone, 2015), drawing a line between those who are outside and those who are inside.

Groups of people who are inside and generate hate speech function as echo chambers, that are named also filter bubbles (Kilvington, 2021). Even actions as liking and following posts of individuals and groups with similar interests on social media, transform these bubbles in grounds for division and radicalization (William et al., 2019). Indeed, ideas propagated within these spaces create misinformation and are fed by confirmation bias; at the same time, the mechanism of ranking algorithms contribute to nurture and reinforce these filter bubbles.

Nevertheless, if digital spontaneism of individuals within filter bubbles and the algorithms of social media are noteworthy, the new phenomenon of so-called hate contractors is growing up, making the propagation of online hate speech even more difficult to comprehend and to contrast. As De Vita

(2020) affirms, recent studies in Europe and in the United States have highlighted that there are individuals who receive compensation for the dissemination of propaganda posts with historical falsifications, through which an idea becomes authoritative. According to a recent classification of Wardle & Derakhshan (2017), information conveyed through these posts is mis-information if information is false, but is not created with the intention to cause harm. Dis-information instead is false and deliberately created to harm a person or a social group. The third kind of information, mal-information is that characterizing hate speech and is similar to disinformation for the intention to inflict harm; what differentiates it from the previous is that mal-information is based on reality.

This kind of pre-constituted material is organized by the centers of hate, which have a defined structure and precise roles. Hate influencers have certainly a crucial and direct role, but there are also other propagators like repeaters (De Vita, 2020), that multiply and amplify hate speech in a capillary way. Two models explain the escalation of hate and the desensitization of those individuals and groups that manifest such behaviors. The most common model is the Pyramid of Hate (ADL Education Division, 2018), that show biased behaviors growing in complexity from the bottom to the top, where the lower level support the upper ones. Thus, as anybody moves up this pyramid, behaviors have worst consequences. Nevertheless, if institutions consider behaviors at lower levels as acceptable, behaviors at the upper levels becomes ever more accepted. The other model to describe this phenomenon is that of Virtual Stages of Hate (Kilvington, 2021), which revises goffmanian model identifying factors that motivate online hate speech. According to this model, online communication is generated within a blurred space; thus, boundaries between backstage and frontstage are mixed up and, in this sense, these conditions represent the ideal context for researchers, within which they are able to observe ideologies and behaviors that once were reserved for the backstage. Once factors provoking online hate are well known, it could be possible to intervene at an educational level to prevent and contrast this widespread phenomenon. In this sense, media education can operate towards it through the promotion of formal and informal learning initiatives targeted at children and youngsters to make them critically digital literate. As Santerini (2019) affirms, in the face of hate speech that pollutes, wounds and hits innocent and defenceless human targets, the need for an educational ecology of the web arises.

3. Media ecology perspectives

Ecological approaches could be used as a basis for theorising educational strategies to prevent online hate speech. Such approaches, supporting the comprehension of children's online experience, consider the media environment as an interplay between technology and society (O'Neill, 2015), where communication shape their behaviours, their social life and their growth. The impact of social media on young persons is noteworthy and the ambivalence embedded within them cannot be ignored: the social media exposure increase risks and could have a dangerous impact on children and youngsters; nevertheless, their potentialities, from an educational, social and cultural point of view, are essential for human development. Notable media ecologists like Postman, McLuhan and Ross (2009) have provided empirical evidence of the educational potential of media technologies.

In the context of the fight against online hate speech, the environment itself of social media, which is the realm where this phenomenon occurs, could be a privileged setting to use for educational

initiatives, even if education is not achieved only online and, according to a media ecology perspective, involves the whole community that revolves around a student.

It is within this context that another ecological approach was taken in consideration for this study. The so-called bio-ecological model developed by Brofenbrenner, created in the 1970s (Brofenbrenner, 1979) and further developed in the following decades (Brofenbrenner, 2005), is a model that defines human development as a process of reciprocal interaction between a child and his environment in which he lives. The model is depicted in the form of concentric and nested circles, which represent spheres of influences of increasing complexity. In this study, the ecological model of Brofenbrenner was updated considering the educational field and the specific contribution of media education.

4. A framework on educational initiatives to prevent online hate speech

Recovering the well-known model ecological by Bronfenbrenner in a media-educational key, the young student is placed at the center of the model, whose education - aimed at making him comprehend and prevent online hate speech – is influenced by the four concentric subsystems (micro, meso, eso and macro) and their interconnections. If we assume an ecological and systemic perspective, learning can therefore be considered as a process that arises from the interaction of individual, social and cultural factors traceable through these four structures (Repetto, 2020). The entities belonging to these systems - the school, the family and the community - interact and synergically collaborate, taking into account student's educational and social needs. Thus, the student can benefit from the supportive relationships that these entities establish with her/him, becoming more digitally competent, able to communicate through social media in an adequate manner and to recognise and contrast online hate speech, as also set out in the European Digital Education Action Plan 2021-2027 (EC-European Commission, 2020).

4.1. The concentric structure of the framework

The four concentric subsystems represented in the left chart of Figure 1 gather the various actors that, at each level, could have a key role in affecting the education of the student. The microsystem is the structure in which the student is embedded, that includes subjects and entities with whom the student has a direct contact such as his/her own family and relatives, the teachers, his/her friends, and classmates. The mesosystem includes interrelations among two or more entities of the microsystem, such as those between school and family or those between friends and family; but at this level, considering the specific topic of online hate speech, other possible entities are youth organisations, advocacy groups or even the social groups or minorities who are discriminated. The relationships with these subjects or groups affect the student and, in turn, may be affected by him/her. The esosystem indicates the broader social system, with which the student does not have a direct relationship, such as institutions, work and social organizations connected with the student, although he/her is influenced by it through the interaction with its microsystem. In the context of social media and hate speech prevention, institutions involved are media organisations, media-regulatory authorities, the Ministry of Education, higher education institutions, NGOs, grant making foundations and the civil society.

The fourth and most external level, the macrosystem, includes the values, ideologies, principles, norms and heritage of the socio-cultural context of reference as well as the life quality, the level of economic development and local politics; it interacts with the structures placed at the other levels, influencing them in a cascade. In the specific context of media education, it concerns human rights regulations, the cultural heritage and educational policies.
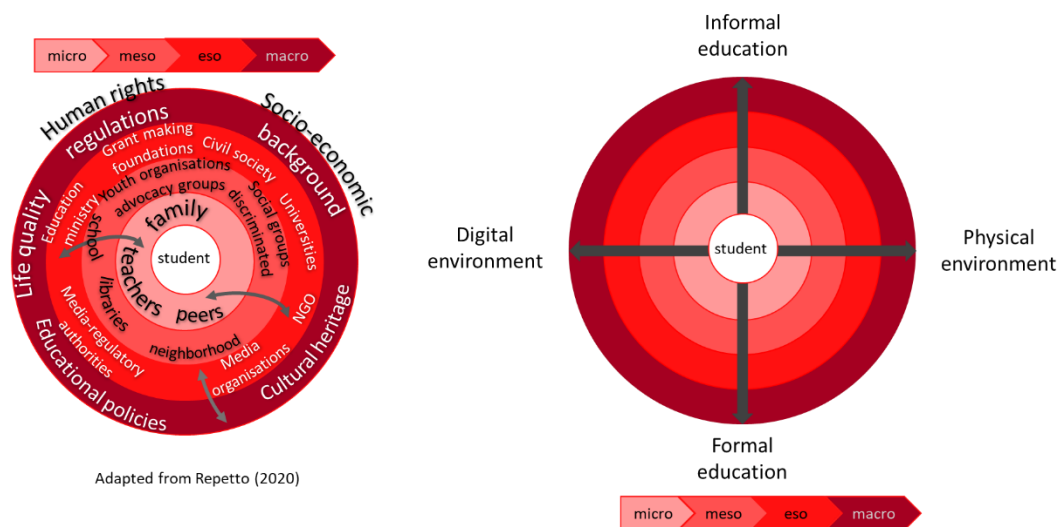

Figure 1: The framework on educational initiatives to prevent online hate speech

4.2. The matrix structure of the framework

A taxonomy has been defined in order to consider the various forms of collaboration and the types of educational projects that can arise from various entities at the described above levels. As exemplified in the right chart of Figure 1, two orthogonal axes have been added to the concentric structure, that represent, respectively, the prevailing learning environment in which media education takes place (horizontal axis) and the educational context in which the initiative is proposed (vertical axis). The horizontal axis brings back to one extreme, the physical environment and, at the other extreme, the digital environment, while the vertical axis presents at the extremes the informal education and the formal one.

Educational projects can occur in a physical place (i.e. in the classroom, in the library or at the headquarters of an association), or could adopt as the main arena an elearning platform, a website, or even the elective place for online communication, the social media. Educational initiatives for prevention of online hate speech are often organised by schools, also in collaboration with associations on the territory; but also the various actors of the local community can promote and entirely managed initiatives of this kind, targeted at children and youngsters.

The two axes divides the structure made up of concentric circular shapes into four quadrants, each of which is therefore delimited by two semi-axes and a series of parallel arches. The areas of this matrix can be used to classify types of educational initiatives that assign to the student, who actively contributes to the construction of the environment in which he lives and whose possible behaviours

have repercussions at the micro, meso, exo or macrosystem level, the role of dynamic protagonist. Huge educational projects that have economic, social or environmental repercussions on the territory or that affect attitudes and values provisions, are placed in the outermost segments; while small initiatives that have a more limited and circumscribed impact are located in the innermost areas.

4.3 A taxonomy of media educational initiatives

The taxonomy developed in this study has been used to classify all the possible media educational initiatives for the prevention of online hate speech and is based on the four quadrants framework described above. This taxonomy was built after an accurate review of the most widespread media educational initiatives and projects dealing with hate speech prevention, that was performed both through a literature review about this research topic and through web documentation to gather online journal articles, reports and project data.

The first quadrant, delimited by the semi-axes informal education / physical environment, includes informal media-educational initiatives promoted by various associations and entities of the local community, which mainly take place in their headquarters. At the micro level of the first quadrant, there are informal and extemporaneous educational activities dealing with the social media, between students and their peers, families or other groups. At the meso level, we may find intercultural initiatives organised in public spaces like in the squares and various kind of learning activities promoted by associations, volunteering and public engagement initiatives centred on the fight against online hate speech. Educational projects promoted by local organisations and funded at a national or European level are located at the macro level.

The second quadrant, delimited by the semi-axes formal education / physical environment, includes the media educational initiatives managed by the schools in their buildings, with the possible support of various entities and organisations of the local community. At a micro level, activities carried out by schools are usually part of curricula and of a unique subject domain; nevertheless, schools organise also interdisciplinary workshops on larger themes such as human rights, peace, media education or cyberbullism, even in collaboration with local organisations, in which hate speech represents a specific topic. At a meso level, schools lead service-learning initiatives that involve students providing them with opportunities to develop civic engagement skills. The main outcomes of these initiatives are the activation of specific services, which constitute a common good as well as a virtuous model of active citizenship, such as listening centres for students or other persons of the local community who are victims of online hate speech. At the same level, initiatives of apprenticeship promoted by school in key workplaces may be considered another opportunity, for students, to comprehend and deal with issues related to discrimination. At a macro level, we may find national educational projects on online hate speech prevention funded by the Ministry of Education or by EU Commission, managed by schools. These projects aim at raising awareness and social intervention; these may affect local policies, as well as modify attitudes and values of the reference population. The initiatives placed in this second quadrant make the school a centre of social and cultural aggregation, in which also local community is involved by contributing as expert - individually or organized in associations, including also persons belonging to discriminated minorities - to the educational activities of the school, or by participating as users and recipients of educational projects.

The third quadrant, circumscribed by the formal education/ digital environment half-shafts, includes the media educational initiatives promoted by the school in collaboration with local entities, mainly managed in environments accessible from the web. These environments break down the spatial and temporal barriers of the school building while, on the educational level, they can amplify the learning processes on a cognitive and social level. The physical structures previously listed for the second quadrant find a potential counterpart in their virtual transposition, with a more or less extensive impact depending on the level of citizen involvement. Therefore, this area includes, at a micro level, environments such as virtual classrooms for carrying out curricular and extracurricular activities on online hate speech prevention. The area at a meso level involves local organisations that transform service-learning into online initiatives that can lead to the creation of media-educational artifacts; while, at a macro level, there are projects EU or nationally funded such as large initiatives of digital citizenship, media education or school twinning. The use of the web also multiplies the possible interlocutors, who do not belong only to the local community, but who can also be located in other cities or abroad. This is the case of the virtual environments, in which twinned classes of different countries collaborate in an active citizenship project. The fourth quadrant is demarcated by the informal education / digital environment semi-axes and is the one in which the offer of informal educational initiatives dramatically multiply, since the resources that the student can draw on are not limited to those offered by the local community. Thanks to technologies, these open and online resources can in fact also be made available by entities located everywhere. Examples of initiatives that fall into this quadrant are the most varied. At the micro-level, personal and group media production or counter-actions such as counter-narratives (Botto, 2019a, 2019b; Pasta, 2018) as main strategies to fight online hate speech, are noteworthy instances. Initiatives that belong to the meso level are those online projects promoted by libraries, the digital labs managed by local organisations fighting for human rights (i.e. virtual tours on the life of minorities) and projects on media production promoted by universities or other entities. At a macro-level, large funded projects like awareness campaign on human rights find their place or other projects with a high civic and social value involving online communities whose active participants countering hatred become real advocates.

5. Conclusion

The issue of online hate speech is of great concern nowadays and research in this field requires a systemic and interdisciplinary contribution for actions aimed at preventing and contrasting this widespread phenomenon. From an educational point of view, media educational initiatives should be promoted and be pedagogically founded to be effective, taking into account also sociological, legal and psychological implications. The aim of awareness-raising actions to face online hate cannot be limited to recognising it, but should tend to the development of judicious critical thinking skills (Santerini, 2019) that could lead to participatory and solidarity citizenship (Bornatici & Vacchelli, 2021). Nevertheless, as underlined by Ranieri and Fabbro (2016), it's difficult raising awareness among young people if proposed teaching and learning activities are far from people interests and concerns. Thus, teachers and educators play a key role to motivate and actively involve students and young people in media educational activities; nevertheless, they must in their turn be adequately trained to become critically digital literate and able to design and realise educational initiatives that transform students in critical and active citizens in the digital realm as well as in the physical reality.

The framework described in this contribution define, through a taxonomic model, some scenarios on the possible educational initiatives that arise from the alliance between the key organisations of the local community to which students belong. The systemic vision adopted in this framework places the student at the center, consistently with the priority of the work plan themes discussed in the Bologna Process (Klemenčič, 2019), which tackled aspects that helped make learner-centred educational approaches a reality. Moreover, this framework considers the interconnections between the various social structures involved in education, formal or informal, giving them a propulsive role. In this setting, each student can take the opportunity to become an active member and to live meaningful learning experiences, while perceiving to be useful and part of a cohesive and vital community. Prevention and contrast of online hate speech could be even more effective if also discriminated people and groups are involved in these educational initiatives, because students should assume the perspectives of discriminated persons and groups, becoming more aware and competent to struggle this phenomenon and to promote inclusive perspectives.

Social media and digital technologies have a key role in the promotion of educational initiatives captured from this framework. These technologies should be perceived as totally connected with the methodological approaches employed, to the extent that become inherent in the same learning environment that these help to define; these technologies, if supported by adequate teaching and learning strategies, can become an empowering element, not only instrumental but structural, making an essential contribution to the enhancement of educational processes needed to face and struggle online hate speech. This could really be the mission of a new and sustainable approach for media education and represent the distinguishing element with respect to the previous forms of media education.

References:

ADL Education Division: *Pyramid of Hate*. https://www.adl.org/sites/default/files/pyramid-of-hate-web-english_1.pdf (last accessed 12.20.2022).

Bagnato, K. (2020). Online hate speech: responsabilità pedagogico-educative. *Annali online della Didattica e della Formazione Docente*, *12*(20), 195-211.

Bornatici, S., & Vacchelli, O. (2021). CIVES - Citizens Towards Sustainability. Emblematic results of a research. *Form@re - Open Journal Per La Formazione in Rete*, *21*(2), 141-150. https://doi.org/10.36253/form-11346

Botto, M. (2019a), Contrastare le discriminazioni con le contronarrazioni. In M. Talarico (Ed.), *Spunti per educare nell'era del digitale*. Roma: Aracne.

Botto, M. (2019b), Narratives against narratives. A Grounded and Narrative-Based Analysis of Discriminations. In AA.VV. (2019), *Proceedings of the 1st International Conference of the Journal Scuola Democratica*, vol. III (pp. 58-63).

Bronfenbrenner, U. (2005). Growing chaos in the lives of children, youth, and families: How can we turn it around? In U. Bronfenbrenner (Ed.), *Making human beings human: Bio-ecological perspectives on human development* (pp. 185–197). Thousand Oaks: Sage.

De Vita, R. *Gli Hate speech online, quando la comunicazione dell'odio è organizzata*. https://www.leurispes.it/gli-hate-speech-online-quando-la-comunicazione-dellodio-e-organizzata/ (last accessed 12.20.2022).

EC-European Commission. (2020). Digital Education Action Plan 2021-2027: Resetting education and training for the digital age. https://education.ec.europa.eu/focus-topics/digital-education/action-plan (last accessed 12.20.2022).

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

Jubany, O., Roiha, M. (2016). *Backgrounds, experiences and responses to online hate speech: a comparative cross-country analysis*. Barcelona: Universitat de Barcelona.

Kilvington, D. (2021). The virtual stages of hate: Using Goffman's work to conceptualise the motivations for online hate. *Media, Culture & Society*, *43*(2), 256–272.

Klemenčič, M. (2019). *Successful design of Student-Centered Learning and instruction (SCLI) Ecosystems in the European higher education area*. A keynote at the XX Anniversary of the Bologna Process.

Morandini, M. C., Thum-Thysen, A., & Vandeplas, A. (2020). *Facing the Digital Transformation: are Digital Skills Enough?* (No. 054). Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

O'Neill, B. (2015). Ecological perspectives and children's use of the Internet: exploring micro to macro level analysis. *Estonian Journal of Education*, *3*(2), 32–53.

Pasta S. (2018). Razzismi 2.0. *Analisi socio-educativa dell'odio online*. Brescia: Scholé.

Ranieri, M., & Fabbro, F. (2016). Questioning discrimination through critical media literacy. Findings from seven European countries. *European Educational Research Journal*, *15*(4), 462–479.

Repetto, M. (2020). La scuola come polo educativo e socio-culturale nell'era pandemica del COVID-19. *Media Education*, *11*(2), 49–59.

Ross, S. M. (2009). Postman, Media Ecology, and Education: From Teaching as a Subversive Activity through Amusing Ourselves to Death to Technopoly. *Review of Communication*, *9*(2), 146–156.

Sala, A., Punie, Y., Garkov, V. and Cabrera Giraldez, M., *LifeComp: The European Framework for Personal, Social and Learning to Learn Key Competence*, EUR 30246 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19417-0, doi:10.2760/922681, JRC120911.

Santerini, M. (2019). Discorso d'odio sul web e strategie di contrasto. *MeTis-Mondi educativi. Temi indagini suggestioni*, *9*(2), 51–67.

Udoh-Oshin, G., (2017). *Hate Speech on the Internet: Crime or Free Speech?* Undergraduate Honors College Theses, 9. http://digitalcommons.liu.edu/post_honors_theses/9 (last accessed 12.20.2022).

Vuorikari, R., Kluzer, S. & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens*, EUR 31006 EN, Publications Office of the European Union, Luxembourg. doi:10.2760/115376, JRC128415.

Waldron, J. (2012). *The Harm in Hate Speech*. Harvard University Press.

Wardle, C., Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report*, 27 (pp. 1-107).

Williams, M.L., Burnap, P., Javed, A., et al. (2019). Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. In *Crime and Justice Studies*. Oxford: Oxford University Press.