JOURNAL OF E-LEARNING AND KNOWLEDGE SOCIETY
Vol. 22, No. 1 (2026)

INVITED PAPER

Learning from violence. Hate Literacy as a core competence for contemporary citizenship

Mario Pireddu[a,1]

[a]*University of Tuscia, Department of Humanities. Languages, Literature and Cultural Heritage - Viterbo (Italy)*

(published: 23/5/2026)

Abstract

Public debates on hate speech are often framed in terms of regulation, moderation, or prevention, positioning hostile content primarily as a pathological deviation to be removed from digital environments. This article proposes a different perspective, arguing that hate speech can be approached as a critical object of analysis for contemporary citizenship education. Building on the concept of hate literacy, the paper conceptualizes hostile online discourses as pedagogically relevant artifacts that both reflect and actively shape models of citizenship, participation, and belonging. Rather than interpreting online hostility as an automatic outcome of digital technologies, the article situates hate speech within a dynamic interplay between intentional political actors, platform infrastructures, and hegemonic cultural narratives. From this perspective, hate speech functions as a form of informal civic education, contributing to a hidden curriculum through which norms, hierarchies, and exclusions are learned and normalized. The paper outlines hate literacy as a core civic competence, understood as the ability to critically read, contextualize, and deconstruct hostile discourses by examining their discursive, technological, and political dimensions. By reframing violence and hostility as lenses through which power relations and civic subjectivities can be analyzed, the article advances a pedagogical framework that moves beyond moral condemnation toward critical engagement. The contribution concludes by discussing the implications of this approach for citizenship education in platformed societies.

KEYWORDS: Digital Citizenship, Hate Speech, Literacy, Media, Learning.

DOI
https://doi.org/10.20368/1971-8829/1136377

CITE AS
Pireddu, M. (2026). Learning from violence. Hate Literacy as a core competence for contemporary [Invited paper]. *Journal of e-Learning and Knowledge Society*, *22*(1). https://doi.org/10.20368/1971-8829/1136377

1. Introduction

In recent years, hate speech has become a central concern in public debate, policy agendas, and academic research on digital media. Across institutional and scholarly contexts, the phenomenon is predominantly framed as a problem of regulation, moderation, and prevention: harmful content is to be identified, classified, and removed in order to protect individuals and safeguard public discourse. This orientation is reinforced by the absence of a universally shared legal definition of hate speech under international human rights law, which has encouraged the development of operational frameworks focused on risk assessment, severity, and governance mechanisms (United Nations, 2023; ARTICLE 19 2015; Gagliardone et al., 2015).

While these approaches are necessary, they tend to construct hate speech primarily as a pathological deviation within otherwise healthy communication environments. In doing so, they risk obscuring its broader cultural, political, and pedagogical dimensions. A growing body of research in media studies, platform

[1]Corresponding author - email: mario.pireddu@unitus.it – address: via S. Carlo, 32 - 01100 Viterbo (Italy).

studies, and critical discourse analysis suggests that hostile online discourse cannot be fully understood as an anomaly or merely as "bad content" to be removed. Rather, it emerges from the interaction between platform infrastructures, socio-cultural imaginaries, and strategic political actors, contributing to the shaping of digital publics and forms of participation (Scharfbillig et al., 2026; Castaño-Pulgarín et al., 2021; Gillespie, 2018; Klonick, 2018; Matamoros-Fernández, 2017).

From this perspective, hate speech is not only an object of governance but also a socio-discursive practice through which meanings, identities, and hierarchies are produced and circulated. Critical traditions have long emphasized that the force of harmful speech cannot be reduced to individual intention or isolated utterances, but must be situated within broader relations of power, iterability, and social recognition (Butler, 1997; Waldron, 2014). These insights become even more relevant in platformed environments, where visibility, amplification, and algorithmic mediation reshape the conditions under which discourse is encountered, interpreted, and normalized (UNESCO, 2021; Van Dijck, 2018; Varnelis, 2008).

Building on these premises, this article proposes a shift in perspective: instead of treating hate speech exclusively as a problem to be eliminated, it can be approached as a critical object for contemporary citizenship education. In digital environments, exposure to hostile discourse is not exceptional but structurally embedded in everyday communicative practices, especially among younger users (Hawdon et al., 2017; Wachs et al., 2021). As such, hate speech participates in what can be described as an informal or hidden curriculum, through which individuals learn – often implicitly – what counts as legitimate participation, who belongs, and which forms of exclusion are normalized.

This article develops the concept of *hate literacy* as a core competence for citizenship in platformed societies. Drawing on research in critical media literacy and cyberhate education, hate literacy is understood as a compound capability that includes recognizing hostile discourse across contexts, interpreting it as a socio-technical and political phenomenon, and responding through informed and situated civic practices (Ranieri & Fabbro, 2016; Blaya, 2019; Obermaier et al., 2025). Crucially, this perspective moves beyond a purely defensive or moralizing approach, emphasizing instead the development of interpretive, analytical, and action-oriented competencies.

The argument unfolds as follows. First, the article critically examines dominant approaches to hate speech centered on regulation and moderation, highlighting their limits in addressing the cultural and pedagogical dimensions of the phenomenon. It then situates hate speech within a mediological framework that accounts for the interplay between discourse, platform infrastructures, and political dynamics. Building on this, the paper conceptualizes hate speech as part of a hidden curriculum of platformed societies and introduces hate literacy as a key civic competence. Finally, it outlines a pedagogical framework grounded in the idea of "learning from violence", discussing its implications for citizenship education in contemporary digital environments.

By reframing hate speech as an object of critical engagement rather than solely of prohibition, this contribution aims to expand the scope of media and citizenship education. In doing so, it aligns with broader efforts to understand digital literacy not only as a set of technical skills, but as a situated and critical practice embedded in complex socio-technical systems and power relations.

2. Beyond Regulation: Limits of Prevailing Approaches to Hate Speech

Contemporary approaches to hate speech are largely structured around the need to regulate, moderate, and contain harmful content within digital environments. This orientation reflects both normative concerns – grounded in the protection of dignity, equality, and social cohesion – and practical challenges related to the scale, speed, and global reach of online communication (O'Regan, 2018; United Nations, 2023). In this context, platforms have progressively assumed the role of key governors of online speech, developing policies, enforcement mechanisms, and technical infrastructures aimed at identifying and managing problematic content (Gillespie, 2018; Klonick, 2018).

Within this dominant paradigm, hate speech is primarily framed as a category of content to be classified and acted upon – whether through removal, demotion, or other forms of visibility control. The increasing sophistication of moderation systems, including automated detection, "reduction" strategies and the policy changes of the different platforms, reinforces a model in which governance is understood as a problem of classification at scale (Douek, 2022; Gillespie, 2022; Roberts, 2019). However, as several strands of research have highlighted, this framing tends to privilege a content-centric and reactive logic, focusing on individual posts or utterances while underplaying the broader socio-cultural and discursive dynamics in which hate speech is embedded. The traditional paradigm of treating hate speech as a localized anomaly – manageable through content moderation, filtering, and account bans – relies on the flawed premise of a stable, fundamentally civil communicative ecosystem. However, the attention economy has decoupled the value of digital discourse from its constructive social function, replacing civic deliberation with engagement criteria such as outrage, polarization, and identity-driven conflict. In this

landscape, respectful debate and hostile rhetoric do not compete on equal ethical grounds; rather, they coexist within the same informational flow and are evaluated solely by their virality and capacity to validate in-group expectations. Hate speech must therefore be reconceptualized not merely as a collection of offensive utterances to be individually removed, but as a property of a socio-technical ecosystem that structurally rewards affective polarization and divisive engagement over social cohesion.

A first limitation concerns the persistent instability of definitions. The absence of a universally accepted legal or conceptual definition of hate speech has led to the proliferation of operational categories that vary across jurisdictions, platforms, and research traditions (ARTICLE 19, 2015; Elliott et al., 2016). While such flexibility allows for context-sensitive interpretation, it also creates significant challenges for both governance and research, often encouraging reductive approaches based on surface features or predefined taxonomies. In computational contexts, for instance, the need for scalable solutions has led to the development of detection systems that rely on annotated datasets and linguistic proxies, which may conflate distinct phenomena such as offensiveness, toxicity, and hate speech (Fortuna & Nunes, 2018).

A second limitation lies in the tendency to individualize and decontextualize harmful discourse. By focusing on discrete acts of expression, regulatory frameworks risk overlooking the ways in which hate speech operates as part of broader discursive formations and cultural repertoires. Critical discourse and cultural studies have shown that hostility is often implicit, coded, and embedded in narratives, humor, or seemingly neutral framings, making it difficult to capture through formal definitions or keyword-based approaches (KhosraviNik & Esposito, 2018; Matamoros-Fernández & Farkas, 2021). Moreover, such discourses are not merely expressive but performative: they contribute to the construction of social boundaries, the normalization of exclusion, and the reproduction of power relations.

A third issue concerns the infrastructural dimension of platform governance. Moderation is not limited to the evaluation of content after publication, but is deeply embedded in the design of platforms, including ranking algorithms, recommendation systems, and interface affordances that shape visibility and interaction (Gorwa, 2019; Helberger et al., 2018). As a result, the circulation and impact of hate speech cannot be understood independently of these socio-technical conditions. Interventions such as demotion or quarantine illustrate how governance increasingly operates through the modulation of visibility rather than through binary decisions of removal (Copland, 2020). Yet, these dynamics often remain opaque to users and underexplored in educational frameworks.

Finally, prevailing approaches tend to position users primarily as subjects to be protected or regulated, rather than as active participants in the construction and transformation of digital publics. While important research has addressed exposure, victimization, and risk factors – especially among younger populations (Hawdon et al., 2017; Wachs et al., 2021) – less attention has been paid to the ways in which individuals learn to interpret, reproduce, or contest hostile discourse within everyday media practices. Even when educational interventions are considered, they are often framed in preventive terms, focusing on awareness and risk reduction rather than on the development of critical and participatory competences (Blaya, 2019).

Taken together, these limitations suggest the need to complement regulatory and technical approaches with a broader analytical and pedagogical perspective. Rather than treating hate speech solely as a governance problem, it becomes necessary to understand it as a complex socio-technical and cultural phenomenon that both reflects and shapes contemporary forms of citizenship. This shift does not imply abandoning regulation, but re-situating it within a wider framework that accounts for meaning-making processes, power relations, and the informal learning dynamics embedded in digital environments. In this sense, moving "beyond regulation" entails not a rejection of governance, but an expansion of the analytical lens: from content to discourse, from isolated acts to mediated practices, and from protection to education.

3. Hate Speech as a Mediated Phenomenon

Moving beyond a purely regulatory understanding of hate speech requires a shift from a content-centered perspective to a relational and mediological one. Rather than being treated as an isolated category of harmful expression, hate speech can be more productively understood as a mediated phenomenon emerging from the interaction between discursive practices, technological infrastructures, and socio-political dynamics.

This perspective challenges deterministic interpretations that implicitly attribute online hostility to the affordances of digital media alone. While features such as anonymity, scalability, and algorithmic amplification undoubtedly shape the conditions of interaction, they do not in themselves generate hate speech. Instead, research in platform and media studies highlights how hostile discourse is co-produced by the interplay of platform design, user practices, and broader cultural and political narratives (Gillespie, 2018; Matamoros-Fernández, 2017). In this sense, platforms do not merely host hate speech but actively participate in structuring its visibility, circulation, and legitimacy.

A key implication of this shift is the need to conceptualize hate speech as part of broader communicative ecologies. Studies on platformed racism and misogyny, for instance, show how

discriminatory discourse is embedded in specific vernaculars, memetic cultures, and affective dynamics that make hostility both recognizable and, at times, socially acceptable within particular communities (De Blasio & Selva, 2024; Walther & Rice, 2024; Bates, 2022; Harrington, 2022; Rachman, 2022; Pacilli, 2020; Bailey, 2021; Sugiura, 2021; EU Commission, 2021; Marwick & Caplan, 2018; Davidson et al., 2017; Massanari, 2017; Jane, 2016; Hochschild, 2016; Wodak, 2015; Waldron, 2014; Hearn, 1998; Gilligan, 1996; Matsuda et al., 1993). These forms of expression often rely on irony, ambiguity, and coded language, complicating attempts at straightforward classification and revealing the limits of approaches based on explicit markers of hate. At the same time, the mediated nature of hate speech must be understood in relation to political communication and processes of collective identity construction. Hostile discourse frequently operates as a tool for boundary-making, contributing to the definition of in-groups and out-groups and to the normalization of exclusionary narratives. As discourse-historical approaches have shown, such dynamics are not confined to marginal spaces but can be integrated into mainstream political communication, where they are reframed, legitimized, and circulated across media systems (Farci, 2025; Wodak, 2015). In platformed environments, these processes are further intensified by the logics of engagement and visibility that reward polarizing and emotionally charged content. Importantly, this mediated perspective also foregrounds the infrastructural dimension of discourse. The circulation of hate speech is shaped not only by what is said, but by how content is ranked, recommended, and connected across networks. Research on platform governance and visibility modulation demonstrates that contemporary forms of control increasingly operate through subtle adjustments in exposure rather than through outright removal (Gillespie, 2022; Gorwa, 2019). As a result, the boundaries between presence and absence, visibility and invisibility, become central to understanding how hate speech gains traction and influence within digital publics. Furthermore, the dynamics of online hate cannot be fully captured without considering their systemic and networked character. Large-scale analyses describe online hate as an adaptive ecology, capable of reorganizing across platforms and communities in response to interventions, rather than being simply eliminated (Johnson et al., 2019). This suggests that governance strategies targeting individual platforms or pieces of content may have limited effects if they do not account for the broader circulation patterns and socio-technical interdependencies that sustain hostile discourse.

Taken together, these perspectives point to the need for a conceptual shift: from viewing hate speech as discrete content to be managed, to understanding it as a dynamic and situated practice embedded in complex media environments. Such a shift has significant implications not only for research, but also for education. If hate speech is mediated, contextual, and relational, then engaging with it requires competences that go beyond recognition and avoidance, encompassing the ability to analyze infrastructures, interpret discursive strategies, and situate communication within broader socio-political processes.

4. Hate Speech and the Hidden Curriculum of Platformed Societies

If hate speech is understood as a mediated and socially embedded phenomenon, a further step becomes possible: recognizing its pedagogical dimension. Rather than being confined to moments of disruption or deviance, hostile discourse can be seen as part of a broader set of informal learning processes through which individuals encounter, interpret, and internalize norms of participation, belonging, and exclusion within digital environments.

This perspective resonates with the notion of *hidden curriculum*, traditionally used in educational research to describe the implicit transmission of values, norms, and social hierarchies beyond formal instruction. In platformed societies, such processes are not limited to institutional settings like schools, but are continuously enacted through everyday media practices. Exposure to online hate – whether as targets, bystanders, or participants – contributes to shaping expectations about what can be said, who can speak, and how conflicts are articulated in public discourse (Hawdon et al., 2017; Wachs et al., 2021).

From this standpoint, hate speech does not simply reflect existing social tensions; it actively participates in the production and normalization of civic imaginaries. Discursive practices of exclusion, dehumanization, and boundary-making function as implicit lessons about social order, teaching audiences to recognize and reproduce distinctions between legitimate and illegitimate members of a community. As critical discourse studies have shown, these processes often operate through repetition, framing, and interdiscursivity, making hostility appear as common sense rather than as explicit aggression (KhosraviNik & Esposito, 2018; Wodak, 2015).

The pedagogical force of hate speech is further amplified by platform dynamics. Visibility mechanisms – such as trending topics, recommendation systems, and engagement metrics – do not merely distribute content but also signal its relevance and legitimacy. In this sense, what gains visibility is not only seen, but also implicitly validated as worthy of attention. Research on platform governance has highlighted how these mechanisms can contribute to the mainstreaming of controversial or polarizing discourse, blurring the boundaries between marginal and dominant narratives (Gillespie, 2018; Helberger et

al., 2018). Importantly, the hidden curriculum of hate speech is not unidirectional or deterministic. Digital environments are also spaces of contestation, where hostile discourse can be challenged, reframed, or resisted through various forms of civic engagement. Practices such as counterspeech, solidarity actions, and collective reporting illustrate how users do not simply absorb norms, but can actively negotiate and transform them (Jane, 2016). However, these practices themselves are learned and conditioned by the same socio-technical environments in which hate circulates, reinforcing the need to understand learning as an ongoing, situated process.

This framing has significant implications for how hate speech is approached within educational contexts. If hostile discourse operates as part of a hidden curriculum, then ignoring or merely suppressing it risks leaving its implicit lessons unexamined and unchallenged. Educational responses that focus exclusively on prohibition or moral condemnation may fail to address the underlying processes through which meanings and norms are constructed and internalized. Instead, engaging with hate speech as a pedagogical object requires making its implicit curriculum visible. This involves analyzing how discourses are constructed, how they circulate within platform infrastructures, and how they contribute to shaping civic subjectivities. In this sense, the goal is not to legitimize or normalize hate, but to render it intelligible as a site of learning – albeit a problematic and often harmful one – through which power relations and forms of citizenship are enacted.

5. Conceptualizing Hate Literacy

The shift from viewing hate speech as an object of regulation to understanding it as a pedagogically relevant phenomenon calls for a corresponding reconfiguration of the competencies required for contemporary citizenship. Within this framework, it is possible to work on the concept of hate literacy as a core civic competence for engaging with hostile discourse in platformed societies. The notion of hate literacy builds on and extends existing traditions of media literacy and critical digital literacy, which have long emphasized the importance of analyzing media representations, understanding power relations, and fostering active participation (De Blasio & Selva, 2024; Ranieri & Fabbro, 2016). However, while these frameworks provide a crucial foundation, they do not always explicitly address the specific challenges posed by hostile, exclusionary, and harmful forms of communication that are increasingly pervasive in digital environments. Hate literacy is proposed here as a more focused construct, oriented toward the interpretation and civic handling of such discourses.

Rather than being conceived as a purely cognitive or technical skill, hate literacy can be understood as a compound and situated capability that integrates multiple dimensions.

First, it involves a *discursive competence*, that is, the ability to recognize and interpret hate speech beyond its most explicit forms. This includes identifying implicit, coded, or ironic expressions of hostility, as well as analyzing the rhetorical and argumentative strategies through which exclusion and dehumanization are constructed (Matamoros-Fernández & Farkas, 2021; KhosraviNik & Esposito, 2018; Costanza-Chock, 2020). Second, hate literacy requires a *socio-technical and infrastructural awareness*. As discussed in previous sections, the visibility and impact of hate speech are shaped by platform architectures, including algorithms, recommendation systems, and moderation practices. Understanding how these infrastructures operate - how content is amplified, demoted, or made visible - is essential for situating hostile discourse within its broader conditions of circulation (Solanes Corella & Hernández Moreno, 2025; Ward, 2024; Douek, 2022; Gillespie, 2018; Davidson et al., 2017). Third, hate literacy encompasses a *political and cultural dimension*, involving the capacity to contextualize hate speech within wider narratives of identity, power, and belonging. Hostile discourse is often embedded in ideological frameworks that normalize exclusion or mobilize affective responses, particularly in relation to race, gender, religion, or nationality. Interpreting these dynamics requires an understanding of how discourse contributes to the construction of civic boundaries and collective imaginaries (Noble, 2018; Van Dijck, 2018; Wodak, 2015; Varnelis, 2008). Finally, hate literacy includes a *civic and action-oriented component*. Beyond recognition and analysis, individuals need to develop repertoires of response that are informed, context-sensitive, and ethically grounded. Research on cyberhate and online participation highlights the importance of practices such as counterspeech, reporting, solidarity, and bystander intervention, which can contribute to reshaping communicative environments (Obermaier et al., 2025; Bailey, 2021; Copland, 2020; Blaya, 2019). In this sense, hate literacy is not only about understanding hate speech, but also about engaging with it as a site of civic practice.

These dimensions should not be interpreted as separate or sequential components, but as interrelated aspects of a broader competence. The ability to act effectively, for instance, depends on the capacity to interpret discourse and to understand the socio-technical conditions in which it circulates. Similarly, critical analysis without the possibility of action risks remaining abstract, while action without analysis may reproduce the very dynamics it seeks to counter. This conceptualization also requires a careful positioning with respect to existing educational approaches. Hate literacy does not

aim to replace broader media literacy frameworks, nor to introduce a new normative layer based on moral condemnation. Rather, it seeks to make explicit a domain of competence that is already implicitly required in everyday digital life but insufficiently addressed in formal education. In doing so, it shifts the focus from protection and avoidance toward critical engagement and situated judgment. At the same time, it is important to acknowledge the limits and tensions inherent in this proposal. Engaging with hate speech as an object of learning raises ethical and pedagogical challenges, including the risk of exposure to harmful content, the potential normalization of hostile discourse, and the difficulty of translating critical awareness into effective action. These issues do not invalidate the concept of hate literacy, but rather highlight the need for careful pedagogical design and contextual sensitivity.

6. Learning from Violence: A Pedagogical Framework

Reframing hate speech as a pedagogically relevant phenomenon entails a significant shift in how violence and hostility are positioned within educational processes. Within this perspective, "learning from violence" refers to a pedagogical orientation that uses instances of hostile discourse as entry points for critical inquiry. The focus is not on the reproduction of harmful content, but on its *deconstruction*: analyzing how it is constructed, how it circulates, and what kinds of social meanings and power relations it encodes. This approach aligns with traditions of critical media literacy that emphasize the importance of interrogating media texts as sites of ideological production and contestation (Rivoltella, 2020, 2001; Ranieri & Fabbro, 2016; Buckingham, 2003, 1996; Silverstone, 1999; Alvarado & Boyd-Barrett, 1992).

A first key dimension of this framework is *discursive analysis as pedagogical practice*. Learners are invited to examine concrete instances of hate speech – texts, images, memes, comments etc. – not in order to classify them as acceptable or unacceptable, but to understand how they work. This includes identifying rhetorical strategies, implicit assumptions, framing devices, and forms of intertextuality that contribute to the production of meaning. Such an approach is particularly relevant in contexts where hostility is not explicit, but embedded in irony, humor, or coded language (KhosraviNik & Esposito, 2018). A second dimension concerns *the analysis of socio-technical conditions of circulation*. Learning from violence requires situating discourse within the infrastructures that shape its visibility and impact. This involves examining how platform mechanisms – such as recommendation systems, engagement metrics, and moderation policies – affect what is seen, amplified, or marginalized. By making these processes visible, learners can develop a more nuanced understanding of how hate speech gains traction and how governance operates beyond simple removal (Gillespie, 2018; Douek, 2022). A third dimension is *the development of interpretive and contextual judgment*. As emphasized in legal and human-rights frameworks, the meaning and impact of hate speech are highly context-dependent, requiring attention to factors such as target groups, historical conditions, and communicative intent (ARTICLE 19, 2015). From an educational standpoint, this translates into the need to cultivate forms of situated judgment that resist both overgeneralization and relativism. Learners are encouraged to navigate ambiguity, assess degrees of harm, and recognize the limits of rigid categorizations (Markham, 2018; Moriggi & Pireddu, 2017). A fourth and crucial dimension involves *civic response and agency*. Learning from violence is not limited to analysis, but extends to the exploration of possible forms of intervention. Research on online hate highlights a range of practices – from counterspeech and bystander intervention to reporting and collective support – that can contribute to reshaping communicative environments (Obermaier et al., 2025; Windisch et al., 2022; Bailey, 2021; Costanza-Chock, 2020). Educational settings can function as spaces where such repertoires are not only discussed but critically evaluated, considering their effectiveness, risks, and ethical implications.

Importantly, this pedagogical framework requires careful design and reflexivity. Working with hostile content raises significant challenges, including the risk of re-exposure to harmful discourse, emotional distress, or unintended normalization. For this reason, educational interventions must be context-sensitive, scaffolded, and attentive to the positionality of learners. The goal is not to immerse students in unmediated exposure, but to create structured environments in which critical distance and collective reflection are possible.

Moreover, learning from violence implies rethinking the role of the educator. Rather than acting solely as a gatekeeper who filters or prohibits content, the educator becomes a mediator who guides interpretation, facilitates dialogue, and supports the development of critical competencies. This shift aligns with broader transformations in media education, where teaching is increasingly oriented toward navigating complexity rather than providing definitive answers. In this sense, the proposed framework does not offer a set of fixed methods, but a pedagogical orientation grounded in the assumption that understanding problematic phenomena is a necessary condition for engaging with them critically. By treating hate speech as a site of inquiry rather than solely as a target of regulation, "learning from violence" opens up new possibilities for citizenship education in platformed societies.

7. Educational Implications for Citizenship in Platformed Environments

The pedagogical framework outlined above has several implications for how citizenship education is conceptualized and practiced in platformed societies. If hate speech is understood as part of a mediated and informal learning environment, and if hate literacy is conceived as a core civic competence, then educational interventions must move beyond reactive and protection-oriented models toward more complex forms of critical engagement.

A first implication concerns the *repositioning of hate speech within educational curricula*. Rather than being addressed episodically – typically in response to specific incidents or as part of broader discussions on online risks – hostile discourse could be integrated as a structural dimension of contemporary media environments. This implies *treating hate speech not as an exceptional breakdown of communication, but as a recurring and patterned phenomenon through which social norms and hierarchies are negotiated*. Such an approach aligns with critical media literacy traditions that emphasize the analysis of representation, power, and participation as central components of citizenship education (Ranieri & Fabbro, 2016). A second implication relates to *the expansion of the scope of digital and media literacy frameworks*. Existing models often focus on skills such as information evaluation, content creation, and online safety. While these dimensions remain essential, they may not sufficiently address the interpretive and civic challenges posed by hostile discourse. Integrating hate literacy into these frameworks requires foregrounding competencies such as discursive analysis, contextual judgment, and understanding of platform governance, thereby extending literacy from a functional to a more explicitly political and cultural domain. A third implication concerns *the role of platforms as pedagogical environments*. As discussed in previous sections, platforms are not neutral channels but infrastructures that shape visibility, interaction, and participation (Maragliano & Pireddu, 2012). From an educational perspective, this means that learning about citizenship increasingly involves understanding how platform governance operates, including moderation policies, algorithmic curation, and visibility management (Markham, 2018; Gillespie, 2018; Gorwa, 2019). Developing such awareness is crucial for enabling informed participation, as users navigate environments where rules are hybrid, dynamic, and often opaque. A fourth dimension involves *the development of civic agency and repertoires of response*. Research on cyberhate and online participation highlights that individuals are not only exposed to hostile discourse but also positioned as potential actors - whether as bystanders, targets, or participants (Obermaier et al., 2025; Wachs et al., 2021). Educational contexts can support the exploration of different forms of engagement, including counterspeech, reporting mechanisms, and practices of solidarity. However, such engagement must be framed critically, taking into account not only effectiveness but also risks, power asymmetries, and the broader socio-technical conditions that shape possible outcomes.

At the same time, these implications highlight the need to *rethink pedagogical roles and environments*. Educators are increasingly required to navigate complex and potentially sensitive materials, balancing critical analysis with ethical responsibility. This involves creating learning settings that allow for the examination of problematic content without reproducing harm, as well as fostering dialogue that acknowledges diverse experiences and positionalities. In this sense, teaching becomes less about transmitting normative rules of "appropriate behavior" and more about facilitating processes of interpretation, reflection, and situated judgment. A further implication concerns *the evaluation and design of educational interventions*. As existing reviews have pointed out, the field of cyberhate education is characterized by a proliferation of initiatives but a relative scarcity of robust evidence regarding their effectiveness (Blaya, 2019; Windisch et al., 2022). Conceptualizing hate literacy as a multidimensional competence opens the possibility of developing more precise frameworks for assessment, including indicators related to interpretive skills, critical awareness of infrastructures, and forms of civic engagement. At the same time, it calls for longitudinal and context-sensitive research capable of capturing the complexity of learning processes in digital environments.

Finally, these considerations suggest that citizenship education in platformed societies cannot be confined to formal educational settings alone. Informal and non-formal contexts – online communities, peer interactions, and everyday media practices – play a central role in shaping how individuals encounter and make sense of hate speech. This reinforces the need for educational approaches that are not only school-based, but also attentive to the broader media ecologies in which learning takes place.

Taken together, these implications point toward a reconfiguration of citizenship education as a practice that engages directly with the complexities of contemporary communication environments.

Integrating hate literacy into this framework does not provide a definitive solution to the problem of online hostility, but offers a way to address it that is consistent with the realities of platformed societies: not by avoiding or simply removing conflict, but by developing the capacities required to understand, interpret, and respond to it.

8. Discussion

The framework proposed in this article - centered on the concepts of hate speech as hidden curriculum and hate literacy as a civic competence - opens up new analytical and pedagogical perspectives. At the same time, it raises a number of theoretical, methodological, and ethical tensions that need to be explicitly addressed.

A first issue concerns the risk of normalization. Reframing hate speech as an object of analysis and learning may be misinterpreted as a form of legitimation, especially in contexts where hostile discourse is already widespread and socially tolerated. The pedagogical approach outlined here attempts to mitigate this risk by emphasizing critical distance, contextualization, and reflexivity. However, the tension remains: engaging with harmful content requires a careful balance between making it visible and avoiding its reproduction as normalized or inevitable. As critical scholarship has shown, repetition and circulation are key mechanisms through which hate becomes embedded in everyday discourse (Walther & Rice, 2024; Hochschild, 2016; Hearn, 1998; Butler, 1997; KhosraviNik & Esposito, 2018).

A second tension relates to exposure and harm. Educational approaches that involve the analysis of hostile content must confront the possibility of re-exposing learners – particularly those belonging to targeted groups – to forms of symbolic violence. This raises important questions about pedagogical design, including the selection of materials, the framing of activities, and the creation of safe and inclusive learning environments. While avoiding exposure altogether may leave implicit norms unchallenged, unstructured exposure risks reinforcing existing inequalities and producing emotional or psychological distress (Farci, 2025; Pireddu, 2022; Citron, 2014).

A third area of tension concerns the translation from competence to action. The concept of hate literacy emphasizes not only interpretive skills but also civic engagement. However, the relationship between awareness, critical understanding, and effective intervention is neither linear nor guaranteed. Empirical research suggests that factors such as perceived efficacy, social norms, and platform affordances significantly influence whether individuals engage in practices such as counterspeech or reporting (Obermaier et al., 2025). This highlights the need to avoid overly optimistic assumptions about the transformative power of literacy alone, and to situate educational efforts within broader socio-technical and institutional conditions.

A fourth issue involves the role of platforms and structural constraints. While the framework foregrounds user agency and civic competence, it must be situated within environments where key aspects of communication are governed by opaque and asymmetrical infrastructures. Platform governance research has shown that visibility, moderation, and enforcement are shaped by institutional logics, economic incentives, and technical constraints that are often beyond users' control. As a result, the scope of individual and collective action is conditioned by factors that literacy alone cannot fully address.

A broader methodological challenge concerns the operationalization and assessment of hate literacy. While the concept is articulated here as a multidimensional competence, translating it into measurable indicators remains an open research problem. Existing studies provide useful starting points – particularly in linking critical digital literacy to bystander intervention – but further work is needed to develop robust, context-sensitive, and longitudinal evaluation frameworks (Blaya, 2019; Windisch et al., 2022).

These tensions do not undermine the proposed framework; rather, they define its conditions of applicability and its limits. They suggest that hate literacy should not be conceived as a self-sufficient solution, but as one component within a broader ecosystem of interventions that includes regulation, platform design, and social action. At the same time, they reinforce the central argument of this article: that understanding and critically engaging with hostile discourse is a necessary, albeit insufficient, condition for addressing the challenges it poses to contemporary citizenship.

9. Conclusion

This article has proposed a shift in the way hate speech is conceptualized within research and educational frameworks. Rather than approaching it exclusively as a problem of regulation, moderation, or prevention, the paper has argued for its consideration as a pedagogically relevant phenomenon – one that both reflects and actively shapes contemporary forms of citizenship in platformed societies.

By situating hate speech within a mediological perspective, the analysis has highlighted its embeddedness in the interplay between discursive practices, platform infrastructures, and socio-political dynamics. From this standpoint, hostile discourse emerges not as an isolated deviation, but as part of broader communicative ecologies through which norms, hierarchies, and boundaries of belonging are constructed and negotiated. In this sense, hate speech can be understood as contributing to a hidden curriculum that informs how individuals learn to participate in digital publics.

Building on this framework, the article has introduced the concept of *hate literacy* as a multidimensional civic competence. Rather than reducing literacy to the recognition or avoidance of harmful content, hate literacy has been articulated as the capacity to critically

interpret, contextualize, and respond to hostile discourse by engaging with its discursive, socio-technical, and political dimensions. This perspective expands existing approaches to media and digital literacy by foregrounding the specific challenges posed by exclusionary and harmful forms of communication.

The pedagogical proposal developed through the notion of "learning from violence" further reinforces this shift. By treating hate speech as an object of critical inquiry, rather than solely as a target of prohibition, the framework emphasizes the importance of developing interpretive judgment, infrastructural awareness, and civic agency. At the same time, the discussion has made clear that such an approach is not without risks and limitations, including issues related to normalization, exposure, and the structural constraints imposed by platform governance.

Taken together, these considerations suggest that addressing hate speech in contemporary societies requires moving beyond dichotomies such as regulation versus education, or protection versus participation. Instead, what emerges is the need for integrated approaches that recognize the complexity of digital environments and the multiple levels – discursive, technological, institutional – at which hostility is produced and managed. From a research perspective, this implies further work on the operationalization and assessment of hate literacy, particularly through longitudinal and context-sensitive studies capable of capturing its impact on civic practices. From an educational standpoint, it calls for the development of pedagogical designs that can engage with problematic content in ways that are critically rigorous, ethically grounded, and responsive to diverse learning contexts.

Ultimately, the contribution of this article lies not in offering a definitive solution to the problem of online hate, but in reframing it as a site of inquiry that is central to understanding and practicing citizenship in platformed societies. In doing so, it invites educators, researchers, and policymakers to reconsider not only how hate speech is governed, but also how it is learned, interpreted, and contested within the everyday experience of digital media.

References

Alvarado M., Boyd-Barrett O., eds. (1992). *Media education: An introduction*, London, British Film Institute, Open University Press.

ARTICLE 19 (2015). *'Hate speech' explained: a toolkit*, London, ARTICLE 19. https://bit.ly/4clMQYh

Bailey, M. (2021). *Misogynoir transformed: Black women's digital resistance*, New York, NYU Press.

Bates, L. (2022). *Men Who Hate Women: From Incels to Pickup Artists: The Truth About Extreme Misogyny and How It Affects Us All*, London, Simon & Schuster.

Blaya, C. (2019). *Cyberhate: A review and content analysis of intervention strategies*, Aggression and Violent Behavior, Volume 45, 2019.

Buckingham, D. (2003). *Media education: Literacy, learning and contemporary culture*. Cambridge, Eng.: Polity Press.

Buckingham, D. (1996). *Critical pedagogy and media education: A theory in search of a practice*. Journal of Curriculum Studies, 28, 627-650.

Butler, J. (1997), *Excitable speech: a politics of the performative*, London, Routledge.

Castaño-Pulgarín, S. A., Suárez-Betancur, N., Vega, L. M. T., & López, H. M. H. (2021). *Internet, social media and online hate speech. Systematic review*. Aggression and Violent Behavior, Vol. 58, 2021.

Citron, D. K. (2014), *Hate crimes in cyberspace*, Cambridge, MA, Harvard University Press.

Copland, S. (2020). *Reddit quarantined: can changing platform affordances reduce hateful material online?*, Internet Policy Review, 9(4).

Costanza-Chock, S. (2020). *Design justice: community-led practices to build the worlds we need*, Cambridge, MIT Press.

Davidson T., Warmsley D., Macy M., & Weber I. (2017). *Automated hate speech detection and the problem of offensive language*, Proceedings of the 11th International Conference on Web and Social Media, 11 (1), 512-515.

De Blasio E., & Selva D. (2024). *Gender and Culture Wars in Italy: A Genealogy of Media Representations*, Cham (CH), Palgrave Macmillan.

Douek, E. (2022). *Content moderation as systems thinking*, Harvard Law Review, Vol. 136, Issue 2.

Elliott C., Chuma W., Gendi Y.E., Marko D., & Patel A. (2016). *Hate Speech, Key concept paper*. Working Paper, MeCoDEM. https://bit.ly/4cgRQNK

European Union Commission, Radicalisation Awareness Network (2021). *Incels: A First Scan of the Phenomenon (in the EU) and its Relevance and Challenges for P/CVE*, Brussel, EU Commission.

Farci, M. (2025). *Quel che resta degli uomini. Sulla mascolinità*, Milano, Nottetempo.

Fortuna, P., & Nunes, S. (2018). *A survey on automatic detection of hate speech in text*, ACM Computing Surveys, 51 (4).

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

Gillespie, T. (2022). *Do not recommend? reduction as a form of content moderation*, Social Media + Society, Volume 8, Issue 3, 07/2022.

Gillespie, T. (2018). *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*, New Haven, CT, Yale University Press

Gilligan, J. (1996). *Violence: Our Deadly Epidemic and Its Causes*, New York, G.P. Putnam's Sons.

Gorwa, R. (2019). *The platform governance triangle: conceptualising the informal regulation of online content*, Internet Policy Review, 8 (2).

Harrington, C. (2022). *Neoliberal Sexual Violence Politics: Toxic Masculinity and #MeToo*, Cham (CH), Palgrave Macmillan.

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). *Exposure to Online Hate in Four Nations: A Cross-National Consideration*. Deviant Behavior, 38(3).

Hearn, J. (1998). *The Violences of Men: How Men Talk about and How Agencies Respond to Men's Violence to Women*, London, SAGE.

Helberger, N., Pierson, J., & Poell, T. (2018), *Governing online platforms: from contested to cooperative responsibility*, The Information Society, 34 (1).

Hochschild, A.R. (2016). *Strangers in Their Own Land: Anger and Mourning on the American Right*, New York, New Press.

Jane, E. A. (2016). *Online misogyny and feminist digilantism*, Continuum: Journal of Media & Cultural Studies, 30 (3).

Johnson, N.F., Leahy, R., Restrepo, N.J., Velasquez, N., Zheng, M., Manrique, P., Devkota, P., & Wuchty, S. (2019). *Hidden resilience and adaptive dynamics of the global online hate ecology*. Nature. 2019 Sep;573.

Klonick, K. (2018). *The new governors: the people, rules, and processes governing online speech*, Harvard Law Review, Vol. 131, Issue 6.

KhosraviNik, M., & Esposito, E. (2018). *Online hate, digital discourse and critique: exploring digitally-mediated discursive practices of gender-based hostility*, Lodz Papers in Pragmatics 14.1 (2018). Special issue on Narrating hostility, challenging hostile narratives.

Maragliano, R., & Pireddu, M. (2012). *Storia e pedagogia nei media*, Roma, Garamond.

Markham, A.N. (2018). *Critical Pedagogy as a Response to Datafication*, Qualitative Inquiry, Vol. 25, Issue 8.

Marwick, A.E., & Caplan, R. (2018). *Drinking male tears: language, the manosphere, and networked harassment*, Feminist Media Studies, 18 (4), 2018, pp. 543-559.

Massanari, A. (2017). *#Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures*. New Media & Society, 19(3).

Matamoros-Fernández, A. (2017), *Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube*. Information, Communication & Society, 20(6).

Matamoros-Fernández, A., & Farkas, J. (2021), *Racism, Hate Speech, and Social Media: A Systematic Review and Critique*. Television & New Media, 22(2).

Matsuda, M. J., Lawrence, C. R. III, Delgado, R., & Crenshaw, K. W. (1993), *Words that wound: critical race theory, assaultive speech, and the First Amendment*, Boulder, Westview Press.

Moriggi, S., & Pireddu, M., (2017), *Vivere e non sapere. Fenomenologia della post-truth tra educazione e comunicazione*, The Future of Science and Ethics, Rivista scientifica del Comitato Etico della Fondazione Umberto Veronesi, Vol. 2, Num. 1, 06/2017, http://bit.ly/2wBBI8r

Noble, S. U. (2018). *Algorithms of oppression: how search engines reinforce racism*, New York, NYU Press.

Obermaier, M., Schmid, U. K., & Rieger, D. (2025), *Empowerment Is Key? How Perceived Political and Critical Digital Media Literacy Explain Direct and Indirect Bystander Intervention in Online Hate Speech*. Social Media + Society, 11(1).

O'Regan, C. (2018). *Hate speech online: an (intractable) contemporary challenge?*, Current Legal Problems, Vol. 71, Issue 1.

Pacilli, M.G.(2020). *Uomini duri. Il lato oscuro della mascolinità*, Bologna, Il Mulino.

Pireddu, M. (2022). *Architetture relazionali, embodiment, co-enaction o apprendimento nel metaverso*, Colazzo S., Maragliano R. (eds), Metaverso e realtà dell'educazione, Roma, Edizioni Studium, 2022.

Rachman, G. (2022). *The Age of the Strongman: How the Cult of the Leader Threatens Democracy Around the World*, New York, Other Press.

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

Rivoltella, P.C. (2020). *Nuovi alfabeti. Educazione e culture nella società post-mediale*, Brescia, Scholè.

Rivoltella, P.C. (2001). *Media Education. Modelli, esperienze, profilo professionale*, Roma, Carocci.

Roberts, S. T. (2019). *Behind the screen: content moderation in the shadows of social media*, New Haven, CT, Yale University Press

Scharfbillig, M., Lewandowsky, S., Altay, S., Van Alstyne, M., Kozyreva, A. et al. (2026). *Fractured reality - How democracy can win the global struggle over the information space*, Publications Office of the European Union, Luxembourg, 2026, https://data.europa.eu/doi/10.2760/9358883, JRC144603.

Silverstone, R. (1999). *Why study the media?*, London, Sage.

Solanes Corella, Á., & Hernández Moreno, N. (2025). *Racism in the digital age: the impact of social media algorithms on public discourse*, The Age of Human Rights Journal, 25, e9603.

Sugiura, L.(2021). *The Incel Rebellion: The Rise of the Manosphere and the Virtual War Against Women*, Bingley, Emerald Publishing Limited.

Varnelis, K. (2008, Ed.). *Networked Publics*, Cambridge, MIT Press.

Van Dijck, J. (2018). *The Platform Society: Public Values in a Connective World*, Oxford, OUP.

Waldron, J. (2014). *The harm in hate speech*, Cambridge, Harvard University Press.

Walther, J. B., & Rice, R. E. (2024, Eds.). *Social processes of online hate*, New York, Routledge.

Ward, C. (2024). *UN Peacekeeping, OHCHR. A conceptual analysis of the overlaps and differences between hate speech, misinformation, and disinformation*, New York, United Nations.

Wachs, S., Costello, M., Wrigh, M. F., Flor, K., Daskalou V., Maziridou E., Kwon Y., Na, E.-Y., Sittichai, R., Biswal, R., Singh, R., Almendros, C., Gámez-Guadix, M., Görzig, A., & Hong, J. S. (2021). *"DNT LET 'EM H8 U!": Applying the routine activity framework to understand cyberhate victimization among adolescents across eight countries.* Computers & Education, 160, 104026.

Windisch, S., Wiedlitzka, S., Olaghere, A., & Jenaway, E. (2022), *Online interventions for reducing hate speech and cyberhate*, Campbell Systematic Reviews, Vol. 18, Issue 2.

Wodak, R. (2015). *The Politics of Fear: What Right-Wing Populist Discourses Mean*, London, SAGE Publications.

United Nations (2023). *What is hate speech?*, https://bit.ly/3OfHYfr

UNESCO (2021). *Media and information literate citizens: think critically, click wisely!*, Paris, Unesco.