

Hate Speech, Deplatforming, and Cross-Platform Migration in Digital Ecosystems



solidarity with
OTHERS
September 2026

This report has been published by Solidarity With OTHERS. The views expressed herein are those of the author and do not necessarily reflect the views of any donors or affiliated institutions.

Hate Speech, Deplatforming, and Cross-Platform Migration in Digital Ecosystems

Author: Genta Abdiji

© 2026 **Solidarity With OTHERS**

Weiveldlaan 41 Unit D, 1930 Zaventem BELGIUM
www.solidaritywithothers.com
info@solidaritywithothers.com

Solidarity With OTHERS, an international human rights organisation registered in accordance with the Belgian Code of Companies and Associations (**OTHERS AISBL**, BCE number: 0715.742.907).

This publication may be used for advocacy, education, and research purposes, provided the source is acknowledged.

Contents

- 1.** Introduction 2
- 2.** Conceptual Framework 4
 - 2.1. Online Hate Speech Definition 4
 - 2.2. Platform Governance and Moderation 4
 - 2.3. Deplatforming as Visibility Control 5
 - 2.4. Cross-Platform Migration 5
- 3.** Policy Challenges and Analysis..... 7
 - 3.1. European Regulatory Model..... 7
 - 3.2. Visibility Versus Systemic Harm..... 7
 - 3.3. Limits of Platform-Centric Governance..... 8
 - 3.4. Fragmentation of Digital Ecosystems..... 8
- 4.** Policy Options..... 10
 - 4.1. Strengthened Self-Regulation..... 10
 - 4.2. Ecosystem-Level Governance Model 11
- 5.** Policy Recommendations 13
 - 5.1. Adopting an Ecosystem-Based Regulatory Framework 13
 - 5.2. Establishing Cross-Platform Coordination Infrastructure 13
 - 5.3. Strengthening Oversight of Media Amplification..... 14
 - 5.4. Increasing Transparency in Political Communication..... 14
- 6.** Conclusion..... 16
- References 17

1 ● Introduction

Online hate speech has become a persistent governance challenge in contemporary digital societies because it circulates through interconnected digital platforms rather than remaining within a single, contained communication environment. Unlike traditional forms of harmful expression, online content can move rapidly across platforms, audiences, and jurisdictions, reaching large-scale visibility within short timeframes.

Platforms such as Facebook, X, YouTube, and Telegram function as key infrastructures of public communication. They do not only host user-generated content but also actively shape its visibility through algorithmic recommendation channels, ranking mechanisms, and moderation policies. These design choices influence which content is amplified, which is downranked, and which is removed entirely. In response to harmful content, platforms apply a range of content moderation tools, including automated detection systems, human moderation, user reporting mechanisms, and enforcement actions such as content removal, account suspension, and permanent bans. Among these tools, deplatforming has become one of the most widely used interventions.

Deplatforming refers to the removal of users or communities for violating platform rules. It is intended to reduce problematic activity by limiting access to audiences and disrupting communication channels within a specific platform environment. In practice, its effectiveness is often assessed at the level of individual platforms, focusing on whether specific accounts or communities are removed or whether malicious material is reduced within that environment. However, evidence suggests that enforcement at the platform level does not always eliminate harmful activity. In some cases, users and communities removed from mainstream platforms have reappeared in alternative digital environments with different moderation standards, audience compositions, or levels of anonymity. For example, communities displaced from platforms such as X or Facebook have in some cases migrated to Telegram channels or smaller fringe platforms where enforcement is less consistent.

This report distinguishes between cross-platform migration and ecosystem-level governance. Migration refers to the movement of users, communities, and content between platforms following enforcement actions. Online hate speech refers to expressions in digital environments that attack, devalue, or discriminate against individuals or groups based on characteristics such as ethnicity, religion, nationality, gender, or sexual orientation.

In digital environments, abusive material is shaped by both individual user behavior and the structural conditions of platform environments. Platforms do not simply host communication; they actively shape how content circulates through recommendation structures, ranking mechanisms, and engagement-based metrics that determine visibility. These systems can amplify content that generates strong emotional responses, such as anger or fear, because such reactions often translate into higher engagement. As a result, polarizing or inflammatory content

may achieve disproportionate reach compared to its original audience size, even when it originates from marginal or niche communities. Digital communication environments also enable rapid replication, reinterpretation, and redistribution of content across platforms. Once content is published, it can be reshared, reframed, and recontextualized in different environments, making it difficult to contain within a single platform boundary. This contributes to persistence even after original removal.

Ecosystem-level governance refers to regulatory approaches that account for the interdependence of platforms within a broader digital communication structure, where actions in one environment can influence activity in another. The central question of this report is whether deplatforming reduces online hate speech or primarily redistributes it across the wider digital ecosystem.

2. Conceptual Framework

2.1. Online Hate Speech Definition

Online hate speech refers to expressions in digital environments that attack, devalue, or discriminate against individuals or groups based on characteristics such as ethnicity, religion, nationality, gender, or sexual orientation. In digital environments, harmful content is shaped not only by individual user behavior but also by structural platform conditions. Platforms do not simply host communication; they shape how content circulates through recommendation mechanisms, ranking mechanisms, and engagement-based metrics that determine visibility.

Research on alternative digital environments (“alt-tech” platforms) further indicates that the characteristics of the communication space influence the spread of prohibited content. Environments with weaker moderation and higher ideological homogeneity can facilitate the circulation of misinformation and extremist narratives, partly because reduced exposure to counter-arguments lowers critical scrutiny of content.¹ Evidence from studies of far-right online networks also shows that harmful actors often operate across interconnected platforms rather than within a single isolated space. These networks can reorganize and persist after enforcement actions, particularly when migration to alternative platforms enables continuity of communication structures.² This suggests that online hate speech should be understood as a system-level phenomenon shaped by the interaction between user behavior, platform design, and broader communication environments.

2.2. Platform Governance and Moderation

Platform governance refers to the set of rules, policies, and enforcement structures through which digital platforms regulate user behavior and content.

Content moderation includes automated detection mechanisms, human review processes, user reporting mechanisms, and enforcement actions such as content removal, account suspension, and permanent bans. These channels operate at scale and must process large volumes of user-generated content in real time.

However, moderation operates under structural tensions. Platforms must balance legal compliance obligations, user safety, freedom of expression concerns, and engagement-driven business models. These competing pressures shape how rules are interpreted and enforced in

¹ Bär, C., et al. (2023). Study on free-speech platforms and migration dynamics. ACM Digital Library. <https://dl.acm.org/doi/10.1145/3587094>

² Urman, A., & Katz, S. (2020). Digital migration and far-right communities: Telegram and platform displacement dynamics. *Information, Communication & Society*.

practice. As a result, moderation outcomes are often inconsistent. Similar content may be treated differently depending on linguistic context, political sensitivity, or regional enforcement capacity. This leads to uneven policy outcomes across platforms and across different user communities. Research on alternative platforms also shows that governance differences across digital environments strongly shape content dynamics. Platforms with weaker moderation frameworks tend to host higher concentrations of controversial or extremist material, while mainstream platforms typically enforce stricter content rules, resulting in divergent communication environments across the ecosystem.³

Platform oversight therefore functions as a form of private regulatory control over public communication, where platforms define the boundaries of visibility, amplification, and removal.

2.3. Deplatforming as Visibility Control

Deplatforming refers to the removal of users, groups, or communities from digital platforms due to violations of platform policies. Although often framed as exclusion or banning, it primarily functions as a tool of visibility control.

By removing accounts or communities, platforms reduce reach, disrupt communication channels, and limit algorithmic amplification within a specific platform environment. The underlying assumption is that reducing visibility will reduce harmful activity. However, empirical evidence suggests that this assumption is incomplete. Studies of far-right actors show that when accounts are removed from mainstream platforms, network activity does not necessarily disappear. Instead, it may shift to alternative platforms where moderation is weaker and enforcement is less consistent. For example, research on alternative communication infrastructures shows that extremist or politically radicalized actors can migrate to platforms such as Telegram after bans on mainstream services. In these environments, communication networks are often reconstructed quickly, allowing previously dispersed actors to reconnect and maintain coordination structures.⁴

In some cases, these networks may even become more cohesive after migration, as users concentrate in fewer platforms with more permissive moderation environments. This suggests that deplatforming may reshape the structure and distribution of activity.

2.4. Cross-Platform Migration

Cross-platform migration refers to the movement of users, communities, and content across digital platforms following enforcement actions or changes in moderation regimes. Migration is shaped by differences between platforms, including moderation strictness, anonymity levels, technological affordances, audience composition, and ideological alignment. These factors influence where users continue communication after enforcement actions. Empirical research

³ Bär, C., et al. (2023). Study on free-speech platforms and migration dynamics. ACM Digital Library. <https://dl.acm.org/doi/10.1145/3587094>

⁴ Urman, A., & Katz, S. (2020). Digital migration and far-right communities: Telegram and platform displacement dynamics. *Information, Communication & Society*.

demonstrates that migration often flows from mainstream platforms toward alternative platforms with weaker moderation standards or higher tolerance for controversial content. This enables continuity of communication even after removal from larger platforms.⁵ Importantly, migration does not necessarily result in fragmentation or weakening of communities. Instead, social ties, ideological narratives, and communication structures can persist across platforms. Studies show that communities can reorganize after displacement while maintaining internal cohesion and shared identity structures.⁶ In some cases, migration may even intensify connectivity within certain groups, as users consolidate into more ideologically homogeneous environments where dissenting perspectives are less present. This indicates that migration can preserve or reshape networks rather than dissolve them.

Ecosystem-level governance conceptualises digital platforms as interconnected components of a broader communication system rather than isolated regulatory environments. Within this infrastructure, content and communities move across platforms through migration, interaction, and amplification processes. Actions taken on one platform can therefore influence behavior on others, including redistribution, concentration, or reorganization of networks. Evidence from studies of far-right online ecosystems shows that enforcement actions on mainstream platforms can lead to rapid reconfiguration of networks in alternative environments. Rather than disappearing, communities often rebuild connections in new spaces, maintaining continuity of communication while adapting to different governance conditions.⁷ At the same time, research on alternative platforms highlights that weaker moderation and ideological homogeneity can create conditions where misinformation and extremist narratives circulate more easily.⁸ This reinforces the idea that governance conditions across platforms interact to shape overall outcomes. This creates a structural mismatch between platform-specific regulatory tools and the interconnected nature of digital communication. Network-based platform rules and policies therefore focus on regulating flows of content, patterns of migration, and interdependencies between platforms rather than treating each platform as an independent regulatory unit. It shifts analytical focus from isolated enforcement outcomes toward system-wide persistence, adaptation, and circulation across digital communication infrastructures.

⁵ Urman, A., & Katz, S. (2020). Digital migration and far-right communities: Telegram and platform displacement dynamics. *Information, Communication & Society*.

⁶ Bär, C., et al. (2023). Study on free-speech platforms and migration dynamics. ACM Digital Library. <https://dl.acm.org/doi/10.1145/3587094>

⁷ Urman, A., & Katz, S. (2020). Digital migration and far-right communities: Telegram and platform displacement dynamics. *Information, Communication & Society*.

⁸ Bär, C., et al. (2023). Study on free-speech platforms and migration dynamics. ACM Digital Library. <https://dl.acm.org/doi/10.1145/3587094>

3. Policy Challenges and Analysis

3.1. European Regulatory Model

The European regulatory framework for online harmful content is based on a hybrid administrative structure model combining platform self-regulation with supranational legal oversight. Within this model, digital platforms are assigned primary responsibility for content moderation, while regulatory authorities establish baseline legal and procedural standards. In practice, this includes instruments such as transparency obligations, risk assessment requirements, and enforcement cooperation mechanisms.

However, implementation remains inconsistent because compliance depends on platform-specific moderation systems, internal resources, and differing interpretations of legal obligations. The SEENPM risk assessment on North Macedonia highlights that even where formal alignment with European regulatory frameworks exists, enforcement capacity and institutional independence remain key challenges. Weak or fragmented institutional structures can limit the effectiveness of digital regulation, even when legal transposition is formally achieved.⁹ This suggests that regulatory effectiveness is shaped not only by legal frameworks, but also by institutional capacity and governance quality within national and platform-level systems.

3.2. Visibility Versus Systemic Harm

Current regulatory approaches to online harm often rely on measurable platform-level indicators such as content removal rates, moderation speed, and compliance reporting. These indicators primarily capture changes in visibility within individual platforms. However, they do not necessarily reflect broader changes in the circulation of hostile digital content across digital platforms. A reduction in visibility on one platform may coincide with continued or even increased activity in other digital environments. Evidence from North Macedonia illustrates this limitation. Harmful content can persist and intensify through cross-platform dynamics, particularly when it is amplified by media reporting practices. In some cases, media outlets reproduce or reframe hate speech in headlines, which unintentionally expands its reach beyond the original platform context. At the same time, SEENPM highlights gaps in monitoring and enforcement capacity within national regulatory mechanisms. These gaps make it difficult to determine whether interventions reduce harm or simply displace it across platforms. As a result,

⁹ SEENPM Network. (2026). North Macedonia: From Formal Progress to Democratic Fragility in Digital Regulation. Southeast European Network for Professionalization of Media.

there is a significant gap between platform-level compliance indicators and broader harm reduction outcomes in terms of actual harm reduction.

3.3. Limits of Platform-Centric Governance

Platform-centric governance is constrained by the operational boundaries of individual digital services. Enforcement tools such as content removal or account suspension operate only within a single platform environment. They do not extend automatically across the wider digital ecosystem. This creates a key limitation when harmful actors migrate between platforms. Even when enforcement is successful in one environment, activity can continue elsewhere under different governance conditions.

The Metamorphosis Foundation analysis shows how harmful narratives can spread across both media and digital environments through amplification mechanisms. In some cases, media reporting reproduces hate speech directly, including in headlines, which increases its visibility and reinforces its circulation.

SEENPM further identifies structural challenges in digital governance, including politicised institutions, weak enforcement capacity, and fragmented oversight mechanisms. These factors reduce the ability of regulators to consistently address cross-platform dynamics. Together, these constraints demonstrate that the platform-focused model is structurally limited in addressing networked forms of harmful communication.

3.4. Fragmentation of Digital Ecosystems

Digital communication environments are increasingly fragmented across mainstream platforms, alternative platforms, and media settings. Each of these environments operates under different regulatory rules, levels of transparency, and enforcement capacities. In fragmented environments, violative content does not remain confined to a single space. Instead, it moves across platforms and media channels, often changing form or context while maintaining its core narrative structure. The Metamorphosis Foundation analysis shows that media landscapes can unintentionally amplify hate speech when sensationalized reporting practices are used. This increases visibility and engagement around targeted harassment, particularly in politically sensitive contexts.

SEENPM further highlights that fragmented regulatory environments, combined with weak enforcement and limited coordination, create significant gaps in oversight. These gaps allow harmful content to circulate across platforms with inconsistent monitoring and uneven responses. As a result, enforcement becomes uneven, and stronger regulation in one environment may lead to displacement rather than reduction of harmful activity. Online hate speech is shaped not only by platform governance but also by upstream political and media processes that shape how content is produced and distributed. The Metamorphosis Foundation analysis highlights that media actors can amplify hate speech through editorial decisions, including the reproduction of discriminatory statements in headlines without context. These practices increase visibility and emotional impact across both traditional and digital media environments. Political polarisation

can further reinforce these dynamics, especially where media spheres lack independence or rely on sensationalist reporting practices.

SEENPM identifies additional structural vulnerabilities in North Macedonia, including weak enforcement capacity, politicized regulatory institutions, and limited safeguards for media independence. These conditions reduce accountability and allow harmful narratives to circulate with limited institutional control. Together, these findings show that online hate speech emerges from the interaction between political communication, media framing, and digital platforms. In such contexts, harmful narratives can move rapidly across outlets, creating reinforcing cycles of amplification.

4 ● Policy Options

4.1. Strengthened Self-Regulation

This policy option builds on the existing European model of platform self-regulation, reinforcing voluntary commitments through stronger implementation expectations, more consistent monitoring, and improved transparency requirements. The core structure remains unchanged: platforms continue to hold primary responsibility for identifying, reviewing, and removing illegal hate speech, while public authorities and the European Commission maintain a coordinating and supervisory role.¹⁰

Operationally, this approach would deepen the commitments already embedded in the EU Code of Conduct on Countering Illegal Hate Speech Online. Platforms would be required to shorten response times for flagged content, expand transparency reporting with more granular breakdowns of removal decisions, and strengthen internal auditing of moderation systems. This would also include expanding trusted flagger networks and improving structured cooperation with civil society organizations and national authorities. Evidence from EU monitoring between 2016 and 2019 demonstrates that this model has produced measurable improvements in moderation efficiency, particularly in review speed and removal rates. However, these improvements are largely procedural and platform-bound. They capture how quickly content is addressed within a single service, but not what happens after enforcement action occurs. Harmful actors and communities can reappear on alternative platforms with different governance standards, which limits the long-term impact of this approach.

A further limitation is structural dependence on platform incentives. Even under strengthened self-regulation, moderation remains shaped by internal business models, resource allocation, and risk management priorities. As a result, enforcement outcomes remain uneven across platforms and contexts. This option therefore improves compliance and responsiveness but does not resolve cross-platform migration or ecosystem-level redistribution of derogatory material.

This policy option shifts governance from a primarily voluntary compliance model toward stronger, more centralized enforcement within the European regulatory framework. It increases the role of public institutions in defining, supervising, and enforcing obligations related to online hate speech across platforms operating in the EU digital space. At the core of this approach is greater legal harmonization.

¹⁰ Council of the European Union. (2016). EU Code of Conduct on Countering Illegal Hate Speech Online. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

This would involve more consistent EU-wide definitions of illegal hate speech, reducing fragmentation between Member States and ensuring more uniform enforcement standards. Platforms would face clearer and more binding obligations, including stricter compliance deadlines, expanded legal liability for repeated failures, and enhanced requirements for risk mitigation. The Digital Services Act (DSA) already represents a partial move in this direction by introducing systemic risk assessments, independent audits, and transparency obligations for very large online platforms. A strengthened centralized model would extend this logic further, increasing the role of EU-level coordination bodies in overseeing enforcement consistency and potentially reinforcing the authority of the European Commission and designated national coordinators.

This model has clear advantages in terms of legal clarity and institutional coherence. It reduces variability in how platforms interpret and apply hate speech rules and strengthens accountability through enforceable legal obligations. It also improves the capacity of regulators to address cross-border digital services that operate across multiple jurisdictions. However, despite these strengths, the model remains largely platform-centric in its operational logic. Enforcement still occurs at the level of individual services, meaning that harmful content removed in one environment can persist or reappear elsewhere. Cross-platform migration and media landscape-level redistribution are not directly addressed. In addition, this approach raises concerns regarding proportionality, freedom of expression safeguards under the European Convention on Human Rights (ECHR), and the risk of regulatory overreach if enforcement becomes overly centralized or politically contested.

4.2. Ecosystem-Level Governance Model

This policy option introduces a structural shift from platform-centred regulation toward platform-network governance of online harm. It is based on the premise that online hate speech does not remain confined to individual platforms but emerges and evolves through interconnected outlets involving multiple platforms, media organizations, and political communication channels. The model reframes the policy framework away from isolated content removal toward oversight of flows, amplification pathways, and migration patterns across digital environments. This includes attention to how communities reorganize after enforcement actions, how narratives spread across platforms, and how visibility is shaped through algorithmic ranking and media amplification.

Empirical evidence supports this systemic perspective. Urman and Katz (2020) show that far-right networks displaced from mainstream platforms can rapidly reconstitute themselves on alternative services such as Telegram, maintaining ideological cohesion and communication structures. This suggests that enforcement at the platform level may displace rather than eliminate harmful activity. A central component of this model is cross-platform coordination infrastructure. This includes interoperable reporting systems capable of tracking patterns of harmful behavior and coordinated activity across platforms, rather than focusing solely on individual content takedowns. Such mechanisms would help identify repeated migration patterns and reduce the ability of actors to evade enforcement by shifting between services.

Implementation could be supported through EU-level structures under the Digital Services Act (DSA), including coordination via national Digital Services Coordinators and oversight by the European Commission. The model also integrates media governance. As shown in the Metamorphosis Foundation analysis of North Macedonia, media outlets can amplify hate speech by reproducing or framing degrading language without sufficient context. This extends the ecosystem governance approach beyond digital platforms to include media amplification dynamics, recognizing their role in intensifying harm.

In addition, the model introduces transparency mechanisms for political communication, particularly high-reach messaging that may contribute to the normalization of exclusionary narratives. Drawing on Bakija, electoral rhetoric is treated as an upstream driver of online harm. The aim is not to restrict legitimate political expression, but to increase visibility and accountability of influential narratives.

Finally, the model includes oversight of algorithmic amplification. Recommendation and ranking spaces can amplify harmful content by prioritizing engagement. Ecosystem-level institutional oversight therefore includes independent auditing of algorithmic settings, systemic risk assessments focused on amplification pathways, and requirements for platforms to evaluate downstream effects of ranking decisions.

While this approach offers a more comprehensive response to online harm, it also raises implementation challenges, including data protection concerns, risks of excessive surveillance, technical interoperability between platforms, and the need to safeguard freedom of expression. These require clear safeguards and proportionality in regulatory design. Overall, the model aligns regulatory design with the structural reality of digital communication systems. It addresses not only content moderation within platforms, but also the broader dynamics of circulation, migration, and amplification across the digital sphere.

5. Policy Recommendations

5.1. Adopting an Ecosystem-Based Regulatory Framework

Regulators should formally shift from platform-focused governance toward an ecosystem-based framework for addressing online hate speech. This shift is necessary because harmful content does not remain contained within single platforms but circulates across interconnected digital infrastructures shaped by platform policies, user behavior, media settings, and political communication.

In operational terms, this means that regulatory success should no longer be assessed primarily through platform-level indicators such as takedown rates or moderation speed. Instead, evaluation frameworks should incorporate system-level indicators, including cross-platform migration patterns, persistence of communities after enforcement actions, and the redistribution of harmful narratives across different digital environments. Evidence from Urman and Katz (2020) demonstrates that when far-right actors are removed from mainstream platforms, they often reconstruct their networks on alternative services such as Telegram, maintaining ideological cohesion and communication capacity. This suggests that enforcement confined to individual platforms risks displacing rather than reducing harmful activity. Accordingly, an ecosystem-based policy framework should be embedded into EU-level regulatory implementation under frameworks such as the Digital Services Act (DSA), ensuring that systemic risk assessments explicitly consider cross-platform dynamics rather than isolated platform behavior.

5.2. Establishing Cross-Platform Coordination Infrastructure

Effective governance of online hate speech requires institutional mechanisms that operate across platforms rather than within them. Cross-platform coordination infrastructure should enable interoperable systems for identifying patterns of harmful conduct that persist across multiple services.

This would include shared classification frameworks for types of harmful conduct, not limited to content-level takedowns but extended to behavioral patterns and network activity. It would also include coordinated reporting systems that allow flagged accounts, groups, or coordinated campaigns to be tracked across platforms where legally and technically feasible. Implementation would require coordination at the EU level, potentially through Digital Services Coordinators and enhanced roles for the European Commission under the DSA framework. Platforms would be required to cooperate in structured information-sharing processes under strict safeguards,

ensuring compliance with data protection standards and proportionality requirements. Importantly, this infrastructure should not function as a centralized surveillance setup but as a distributed coordination mechanism focused on systemic risk mitigation. Its primary purpose would be to reduce enforcement fragmentation and prevent repeated displacement of harmful actors across platforms.

5.3. Strengthening Oversight of Media Amplification

Regulatory frameworks should integrate media environments into governance approaches for online hate speech, recognizing that harmful content is often amplified before it enters or after it circulates within digital platforms. The Metamorphosis Foundation (2024) demonstrates how media outlets in North Macedonia have reproduced hate speech content directly in headlines, often without contextualisation, particularly in politically sensitive cases. This practice increases visibility and emotional intensity of harmful narratives, contributing to wider dissemination across digital environments. Policy responses should therefore focus on strengthening editorial standards, improving accountability for the reproduction of harmful content, and promoting guidelines that discourage uncritical amplification of hate speech. This does not imply restricting journalistic freedom but ensuring responsible framing and contextualisation of sensitive content. In addition, media literacy initiatives should be expanded to improve public understanding of how hate speech circulates across media and platform ecosystems. This is particularly relevant in fragmented media environments where engagement-driven reporting can unintentionally amplify harmful narratives.

5.4. Increasing Transparency in Political Communication

Political communication plays a key upstream role in shaping the production and normalization of online hate speech. As shown by Bakija, political actors in the Western Balkans frequently use exclusionary narratives, particularly against LGBTQ+ communities, which contribute to legitimizing hostility in both offline and online environments.¹¹ To address this, transparency mechanisms should be introduced for high-reach political communication in digital spaces. These mechanisms should focus on disclosure of large-scale political messaging, including paid amplification, coordinated campaigns, and algorithmically boosted content. The objective is not to restrict political expression but to increase accountability and visibility of influential narratives that may contribute to social polarization or the escalation of harmful discourse. This aligns with broader democratic principles under the ECHR, particularly the need to balance freedom of expression with protection against harm and discrimination. Algorithmic structures play a central role in shaping the visibility and spread of online hate speech.

Recommendation engines and ranking systems often prioritize engagement, which can inadvertently amplify polarizing or violative content. Regulators should therefore require platforms to conduct systematic risk assessments of their algorithmic systems, focusing specifically on amplification pathways. These assessments should evaluate how content spreads

¹¹ Bakija, M. (2025). Balkan political discourse 'Encouraging' online hate towards LGBTQ+ people. Balkan Insight.

across user networks, how recommendation systems influence exposure to harmful material, and how engagement-based ranking may contribute to escalation dynamics. Independent auditing mechanisms should be introduced to ensure transparency and accountability in algorithmic design and implementation. Platforms should also be required to incorporate harm mitigation measures into network design, including adjustments to ranking systems where systemic risks are identified. This approach aligns with the risk-based framework of the Digital Services Act (DSA), particularly its emphasis on systemic risks associated with very large online platforms.

Smaller states with limited regulatory and institutional capacity, such as North Macedonia, require tailored implementation strategies within broader ecosystem-based institutional design frameworks. As highlighted in the SEENPM (2026) assessment, formal alignment with EU digital oversight standards does not automatically translate into effective enforcement. Institutional fragmentation, politicisation of regulatory bodies, and limited technical capacity can significantly weaken implementation outcomes.

Policy design should therefore include capacity-building measures for national regulatory authorities, improved coordination with EU-level institutions, and structured support for independent oversight bodies. Particular attention should be given to ensuring regulatory independence, strengthening technical expertise, and improving cross-sector coordination between media, telecommunications, and digital governance institutions. At the same time, implementation strategies should account for the specific vulnerabilities of smaller information environments, where rapid cross-platform migration and strong dependence on global platforms can intensify the spread of harmful content. This makes cross-platform coordination particularly important in such contexts.

6. Conclusion

Online hate speech should be understood as a structural feature of contemporary digital environments rather than a set of isolated incidents occurring within individual platforms. Its production, amplification, and persistence are shaped by the interaction between platform architectures, algorithmic frameworks, media practices, and political communication environments. This interconnected structure limits the effectiveness of policy approaches that focus exclusively on single platforms. The analysis demonstrates that deplatforming and content removal can reduce visibility within specific services, but they do not necessarily reduce the overall presence of harmful actors or narratives. Empirical evidence from Urman and Katz (2020) shows that communities displaced from mainstream platforms frequently reconstruct themselves in alternative environments, maintaining continuity through cross-platform migration. This indicates that enforcement at the platform level often redistributes rather than eliminates harmful activity. Similarly, evidence from the Western Balkans highlights the role of upstream drivers in shaping online hate speech. Bakija shows how public political narratives can normalize exclusionary narratives, particularly against LGBTQ+ communities, while the Metamorphosis Foundation demonstrates how media amplification practices can further intensify and disseminate offensive material. These dynamics illustrate that online hate speech is not solely a platform governance issue but the outcome of broader socio-political and media environments.

The policy implications of these findings are significant. Platform-centric regulatory models, including strengthened self-regulation and centralized enforcement, improve procedural compliance and moderation efficiency but remain insufficient in addressing system-wide harm. They do not adequately capture cross-platform migration, ecosystem fragmentation, or amplification dynamics across interconnected communication environments. As a result, the report supports a shift toward an ecosystem-level regulatory approach. This approach reframes regulatory design around flows of content, movement of communities, and systemic amplification processes rather than isolated platform actions. It emphasises coordination across platforms, integration of media governance considerations, and oversight of algorithmic systems that shape visibility.

The core conclusion is that reducing visibility in one digital environment does not equate to reducing harm in the broader digital ecosystem. Effective institutional oversight requires mechanisms that account for persistence, redistribution, and adaptation across platforms and media systems. Only an ecosystem-based approach can adequately address the structural nature of online hate speech in contemporary digital societies.

References

- Bakija, M. (2025). Balkan political discourse 'Encouraging' online hate towards LGBTQ+ people. Balkan Insight.
- Bär, C., et al. (2023). Study on free-speech platforms and migration dynamics. ACM Digital Library. <https://dl.acm.org/doi/10.1145/3587094>
- Council of the European Union. (2016). EU Code of Conduct on Countering Illegal Hate Speech Online. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en
- European Commission. (2019). JHA Council 7 October – Information note: Progress on combating hate speech online through the EU Code of Conduct (2016–2019). (Council document 12522/19 MMA/ns).
- Metamorphosis Foundation. (2024). From inbox to front page: How the media fuels hate speech in North Macedonia. Global Voices.
- Urman, A., & Katz, S. (2020). Digital migration and far-right communities: Telegram and platform displacement dynamics. Information, Communication & Society.
- SEENPM Network. (2026). North Macedonia: From Formal Progress to Democratic Fragility in Digital Regulation. Southeast European Network for Professionalization of Media.



Solidarity With OTHERS

Belgium, 2026

www.solidaritywithothers.com

info@solidaritywithothers.com