Page 753 - 764 **DOI:** *https://doij.org/10.10000/IJLMH.119123*

INTERNATIONAL JOURNAL OF LAW MANAGEMENT & HUMANITIES

[ISSN 2581-5369]

Volume 8 | Issue 2

2025

Defamation in the Digital Age: Social Media's Role in Amplifying Hate Speech and Challenging Free Expression

ISHAN RANJAN[1]

ABSTRACT

Social media has transformed global communication because most individuals can share their views and contribute to the discussion of issues. However, the ease with which all people share posts on social media has facilitated the spread of hate speech, which incites violence, discrimination, and hostility against vulnerable groups. This paper analyzes the impact of social media on defamation and freedom of expression, raising questions about the rise of social media platforms and the type of function they lend to hate speech. The regulation of hate speech becomes tricky because social media operates across national spaces with varied laws and legal instruments. Although efforts to curtail hate speech on social media are still inadequate, there is still no common regulatory framework for all: whereas some, such as the United States, place the principle of free speech at its core and others, such as the European Union, apply stronger regulation over hate speech. Algorithms on these sites often prefer stimulating and emotionally resonant content and give immense play to inflammatory rhetoric. The paper continues with a contrast between the various legal frameworks used in defamation proceedings across the UK, India, and the US and tensions between free speech and the protection of reputation. It ends with recommendations on how to address hate speech, stressing a more holistic approach that would entail stronger intergovernmental collaboration with technology companies and civil society. The paper is based on a doctrinal methodology of research.

Keywords: *defamation, social media, legislations, hate speech, freedom of expression..*

I. RISE OF SOCIAL MEDIA AND HATE SPEECH

Gone are the days when people communicated, connected with the world, and engaged with each other. With Facebook, Twitter, and YouTube, it has changed the way people communicate, and it makes easy expression to others of thoughts, news sharing, and even joining a conversation about certain issues. Yet, in doing such awesome contributions, social media has "allowed" content detrimental to society, including hate speech, to spread really fast. Defined

[1] Author is a student at Christ University, India.

 [ISSN 2581-5369]

as the type of speech that incites the inflaming of violence, discrimination, or hostilities towards people or groups based on attributes like race, ethnicity, religion, gender, and sexual orientation, hate speech has become grave in these forums. The billions of users daily engaged on social media heighten its ease of dissemination and therefore its increased predominance and impact.

The nature of social media allows hate speech to thrive in ways that cannot be done by traditional media. Social networks provided the global platform where anyone can instantly post content, bypassing some of the scrutiny and filtering usually associated with traditional communication. As a result, hate speech exploded rapidly, was often unchecked, but spread quickly. This anonymity and pseudonymity veiled upon these portals allows users to mask behind false identities, forcing them to spread hate speech messages without facing immediate consequences in terms of conduct. Such anonymity proves to be a hurdle in identifying perpetrators: it encourages the continued dissemination of radical and hateful rhetoric.

This is complicated further by technological realities such as "mirror sites."[2] If offensive speech is banned on one site, it will merely reappear on another site or be hosted in a different form, so it can continue to spread even as efforts are made to ban it. This reflects the shortcomings of existing regulatory approaches, with such approaches usually struggling to keep pace with the rapid evolution of online technologies. Therefore, since hate speech is not ultimately eradicable from the digital domain, then hate speech will remain an issue that continues to evolve over time for the companies and governments who seek to regulate it.

Other algorithms involving social media which would impact improving user engagement value promote provocative material as much as possible.

Therefore, hate speech and much more divisive, hateful content could become even more widespread, have a larger reach to audiences, or even create a sense of normalcy with hate speech. Further exposure to the same users to hateful speech has the potential of such beliefs being institutionalized in particular communities, thus encouraging intolerance and hateful attitude.[3] The institutization of hate speech has more deeper implications on the larger social sphere as it can lead to discrimination and violence against certain groups.

II. LACK OF COMMON DEFINITION OF HATE SPEECH (HIERARCHY OF HATE)

Hate speech is very quickly becoming an organizing theme in regulating online content, at least

[2] Brittan Heller, Of Legal Rights and Moral Wrongs: A Case Study of Internet Defamation, 19 YALE J.L. & FEMINISM 279 (2007).
[3] Fernne Brennan, 'Legislating against Internet Race Hate' (2009) 18 Information & Communications Technology Law 2, 123.

within the context of social media. Despite the growing concern over increased instances of hate speech on the Internet, there remains little, if any, consensus on what hate speech actually is.[4] In fact, the lack of definition-one that could be applicable across the world-is one of the major barriers to speaking cross-border and disparate legal systems regulating harmful speech. Legal frameworks vary in other places, and even web sites have defined their own understandings of the term when it comes to hate speech. The social, political, and cultural context within each varies accordingly so that consistency in the regulation and application of the usage of hate speech is not usually pursued in such concerns.[5] One reason is the problem the term hate speech presents: on the one hand, the need to protect free speech and, on the other, a need to restrain unnecessary negative rhetoric.

While many believe that some form of speech—loosely called hate speech or offending speech—is in fact subject to restriction, the distinction between what may and may not be allowed often is less than clear. This ambiguity stands out particularly in democratic societies where free speech is foremost among the freedoms established. For instance, free speech is seriously protected in the United States by laws embodied in the First Amendment, including hate speech. Many of Europe's countries, victims of Nazi-oriented war and inclusive of the Holocaust, made hate speech statutes more stringent by criminalizing certain levels of racism, xenophobia, and denial of the Holocaust.

Such a stance on free speech versus hate speech makes for this sort of global confusion over how to define and regulate hate speech. The inconsistencies in definitions and regulatory frameworks have driven what some scholars refer to as a "hierarchy of hate." The "hierarchy of hate" concept implies that there are forms of hate speech that more readily take on the name, and therefore the regulation and punishment, such as those based on race and ethnicity. For instance, the European Union's Framework Decision on Combating Racism and Xenophobia lists certain racist and xenophobic speech, while discriminations such as homophobia or ableism are left untreated.[6] Similarly, most conventions through international conventions target the elimination of every form of racial discrimination, such as International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), which mainly targets racially discriminatory activities and excludes other vulnerable groups.

[4] Irene Nemes, 'Regulating Hate Speech in Cyberspace: Issues of Desirability and Efficacy' (2010) 11 Information and Com- munications Technology Law 3,195.

[5] For analysis of the issue of jurisdiction and online hate regulation look at: Natalie Alkiviadou, 'Regulating Internet Hate: A Flying Pig?' (2016) 7 Journal of Intellectual Property, Information Technology and E-Commerce Law 3.

[6] Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law.

In fact, while hate speech against racial or ethnic groups is a grim matter taken seriously, discrimination based on gender, sexual orientation, disability, or religion may never be taken seriously or addressed effectively by individuals and communities. That "hierarchy of hate" speaks to broader social attitudes toward which groups deserve protection from harmful speech. Historical and cultural contexts usually contribute to such a hierarchy wherein violence was always associated with racial discourses. While, however, violence to other types of hate speech is excluded from legal and regulatory frameworks, this exclusion may somewhat reinforce the already established marginalization of vulnerable groups.

If regulatory bodies refuse to use their authority to address hate speech against LGBTQ+ people, religious minorities, and those with disabilities, for example, they may, in some measure, perpetuate a system that deems certain sorts of harm more important than others. Simply put, in the absence of clear singular definition of hate speech, damaging rhetoric gets past regulations and into spaces that garner a clear unified approach. This often creates a hierarchy, with some forms of hate speech being strictly regulated more than others, reflecting societal biases on the groups in the greatest need of protection. Such extensive measures to combat hate speech necessitate alterations to the legal and cultural landscape in order to ensure that all forms of prejudice are viewed with the same gravity, treated fairly, and given the same urgency.

III. CHALLENGES IN REGULATING HATE SPEECH ON SOCIAL MEDIA:

The method to control and regulate hate speech on the social media has become the hot topic of discussion over the internet as it seems to have acquired a central role in the communications sector. People from all parts of the globe can share their ideas and their thoughts and even their opinions. The opened-up environment has fostered engagement and connections but also paved the way for risky rhetoric and hate speech. This is a very dangerous form of social media hate speech, as it will incite violence, encourage discrimination, and emotionally harm people and groups. On the other hand, the global nature of the internet, differing national legal standards and the innate clash between the protection of free speech and prohibiting harm do make regulation of hate speech through those channels simply too complex. Therefore, a general overarching legislative structure significantly is an impediment in regulating hate speech through online media.

Hate speech attracts different strict levels of regulations in different countries. For instance, in the United States, the First Amendment vigorously upholds free speech in circumstances involving offending or damaging speech. This freedom makes it difficult to place restrictions on hate speech. On the other hand, most European countries have stricter rules regarding hate

speech and deem it a criminal activity there under the provisions of racism and xenophobia or as incitement to violence. Social media sites are global in nature, therefore, what could be illegal hate speech in one country, might actually fall under protected speech under another rule of law, making the task even more difficult to set common standards for border-crossing regulations.

Another crucial challenge is the vast quantity of materials that are usually uploaded onto social media platforms. Billions of people around the world post content every day, which indeed makes it utterly impossible for any social media company to monitor such activity in real time. Instead, platforms like Facebook, Twitter, and YouTube rely heavily on user reports to flag harmful content. Because these companies are producing so much content with so many users, they have had to develop user-based reporting systems. This model is inherently reactive. Thus, hateful content usually remains online until it is detected and addressed. Meanwhile, hate speech could be spread to large audiences or even incite harm or even facilitate digital toxicity.

Furthermore, the algorithms controlling engagements on a multitude of platforms offered on social media that incentivize the most provocative content enhance hate speech amplification. The more provocative and emotive a post, then the likelihood that it attracts more comments, likes, and shares. And these interactions can further make the post promoted by the algorithm of the platform. This creates a dilemma because to maximize the engagement of users with the product, platforms have little means to avoid inadvertently fuelling the spread of harmful content, including hate speech.

There is a fine balance between the desire for engagement and the imperative to not let dangerous rhetoric go out into the world.

Furthermore, the enforcement of existing hate speech laws is not uniform due to its reliance on these self-imposed company guidelines. For instance, there are community standards by Facebook and YouTube against hate speech, but enforcement is not uniform. Some users and contents are blocked within minutes of posting, yet others go on to thrive even after repeated complaints. This also partly lies in the fact that platforms fear firmly taking a stance since it will attract protesting users or accusations of censorship.

This is a complex process wherein the regulation of hate speech on social media is hampered by inconsistency in legal systems across various countries, the vast nature of user-generated content, the algorithms used to favour divisive material, and selective enforcement. Such challenges will require a coalition of various stakeholders - closer collaboration among governments, the tech industry, and civil society-at-large in order to pool their efforts into the

 [ISSN 2581-5369]

development of technology that can be even more complex in detecting and preventing hate speech without stomping on rights to free speech.

IV. COMPARISON OF LEGAL FRAMEWORKS

Defamation laws vary from country to country inasmuch as each has its own historical, cultural, and legal traditions. Thus, a particular country may emphasis free speech and freedom of expression, while another country emphasizes protecting individual dignity and reputation. And this is probably the case in social media regulation, wherein the 'balance' of protection against defamatory content and upholding free expression becomes too overwhelming. Over the past few years, countries around the world, such as India, have either adopted or fashioned new legal tools to contend with the threats posed by online defamation. Defamation and Free Speech: Global Perspectives

One of the greatest challenges in the regulation of defamation is the tension between the speech and the reputation. Various countries have responded differently to this issue in terms of constitutional protections, legal precedents, and social values. Free speech is entrenched almost as a right to be let absolutely free under the First Amendment to the Constitution of the United States. This has made the defamation laws in America exceptionally stinging, forcing a relatively high bar for how plaintiffs must prove defamation, especially when public figures are involved. The landmark U.S. The Supreme Court case **New York Times Co. v. Sullivan**[7] established the "actual malice" standard, requiring public figures to prove that defamatory statements were made with knowledge of their falsity or reckless disregard for the truth. This reflects a strong preference for protecting free speech, even at the risk that it allows in the way of circulation for possibly erroneous statements of defamation. In contrast, countries like the United Kingdom focus more on the protection of reputation than liberty.

The United Kingdom changed its laws relating to defamation in 2013. The laws still maintain fewer cases of spurious litigation against publishers without omitting responsibility for adverse or offensive material. One of the salient aspects of the change is that a plaintiff will have to show "serious harm" as a prerequisite for filing a lawsuit. This is an attempt by the UK to find a proper balance between freedom of expression and the right to reputation, but the threshold of defamation proof is not so high as in the United States. Besides that, social media may lead to claims of defamation in the UK as well, and it is possible that certain liability of platforms appears if they don't react upon complaints about undesirable content. The European Union has

[7] 376 U.S. 254 (1964)

 [ISSN 2581-5369]

as well its balance in free speech, especially since it is a regionally diversified member states.

Laws on defamation differ from one member of the EU to another, while there is a generic legal framework provided by the ECHR on free speech and reputation. Article 10 provides for the right of freedom of expression, while Article 8 provides for the right to private life, among these is reputation. The ECtHR has produced a series of decisions on defamation seemingly in the quest for some balance between these two competing rights.[8] Thus, for instance, in social media cases, the ECtHR tends to consider the effect of the statement, the reach of the platform, and the measures taken by the platform to reduce harm- factors that'll go on to influence national approaches to online defamation. Germany's NetzDG, or Network Enforcement Act, passed in 2017, stands as the first national law that requires more social media responsibility towards illegal contents like hate speech, defamation, or incitement to violence.

It targets social networks in Germany with more than two million users, like Facebook, Twitter, or YouTube, and requires that evident illegal content has to be removed within 24 hours after being noted. More complex cases are given up to seven days to review and act. In case of non-compliance, this law carries fines up to €50 million, giving it that sense of seriousness in enforcing it. NetzDG stands out with its transparency mode.

Social media companies must issue bi-annual reports on what was done with complaints. These will include the number of complaints received, pieces of content taken down, and the timeframes of actions taken within the deadlines. This serves the purpose of holding people accountable and making them comply with the law.[9] At its worst, the law has been criticized as potentially causing more over-censorship than is necessary in the undermining of free speech, according to proponents, this is a step that is required to stem the emanating epidemic of harmful content on the Internet. NetzDG has set a global precedence in setting the tone for discussions on content moderation and liability of platforms in legal terms. It has inspired more such regulatory efforts in other countries that are concerned with controlling hate speech and illegal online content. The Indian Legal Framework for Defamation

In India, the approach to defamation arises from the colonial legal heritage along with constitutional protections. Indian law treats defamation both as a civil and a criminal offense. In most of the Western countries, defamation is held to be exclusively a civil wrong. The main

[8] Leandro Silva, Mainack Mondal, Denzil Correa & Fabrício Benevenuto, 'Analyzing the Targets of Hate in Online Social Media' Proceedings of the Tenth International AAAI (Association for the Advancement of Artificial Intelligence) Confer- ence on Web and Social Media (2016) 688.

[9] Times Of India, *Social media cos to file action-taken reports each month*, Times of India (July 1, 2021), https://timesofindia.indiatimes.com/india/social-media-cos-to-file-action-taken-reports-each-month/articleshow/84006987.cms.T

source of criminal code of India is Bhartiya Nyaya Sanhita or the BNS, which encompasses an exhaustive list of criminal offenses as well as punishments corresponding to each. It contains provisions related to the criminal defamation, especially Section 356(1) mentions it. The identical punishment has been provided under Section 356(2) of the Act. According to the above sections, an act created and published which harms someone's reputation has been condemned as a criminal offense. This includes the written and oral defamation that can be in digital form also. That being said, it does leave enough space for the growing concern of online defamation and its rapid spread over social media. Section 356(2) sections cover the term of punishment, which might either be imprisonment, fine, or community service. In India, civil defamation is governed by the law of torts and allows an individual to file a case against someone for monetary damages over false statements that have harmed his reputation. The onus of proof lies with the plaintiff whereby, on a prima facie case, he has to establish that the statement referred to is defamatory and was so published to a third party.

Social media cases generally act as the trigger for greater reputational damage due to the expanded reach of such defamation. Indian courts are now in the process of grappling with cases arising from baseless content on Facebook, Twitter, and WhatsApp, which complicates the sheer virality of the information into legal proceedings. One of the major concerns within the context of legal framework in India relates to intermediary liability, namely liability of online intermediary services such as Facebook and Twitter for third-party user defamatory content. The IT Act of 2000 provides an intermediary with limited liability provided it fulfills the conditions required of it under the Act.

Section 79 of the IT Act provides immunity for intermediaries for third-party information, except where they delete or disable access to such prohibited content within 36 hours after being notified. However, IT Rules 2021 have tightened the screws for social media technologies as there will be grievance officers appointed by every social media entity having additional responsibilities, such as a mechanism to redress complaints on defamatory content. These new regulations reflect rising concerns in India over dangerous content that spreads through social media and the need for more robust regulatory control.

Comparative Approaches: Lessons and Challenges

Comparing the legal frameworks used by countries to regulate defamation on social media reveals many common challenges alongside differences in approach.[10] Free speech and the right

[10] J. Lakshmi Charan & J. Krishna Charan, *A Critical Analysis on Cyber Defamation in India: Laws and Issues in Present Scenario*, 12 Eur. Chem. Bull**.** 192 (2023).

to reputation are universal rights, but relative weight assigned by countries to such rights varies. This balance imposes heavy burdens on the plaintiffs - especially public figures - to prove, beyond a reasonable doubt, that the statements in question were made with "actual malice." Very much reflective of the historic commitment of the country to free speech, even at the expense of individuals' reputation, this is very far from a country like the UK and India, where reputation takes a much more protective approach and allows seeking legal recourse with much less of a burden of proof. This balance between free speech and reputation gets complicated in countries like India, where defamation is a criminal offense, by the likelihood of misuse of defamation laws to suppress dissent. Requiring criminal sanctions for defamation, especially in political or public figures cases, raises the cold chill effect of free expression, mostly on social media platforms where user dissent can be expressed very easily. One of the great challenges across jurisdictions is intermediary liability.

Social media is pan-global, but it operates under greatly different national legal frameworks.

The issue is whether these platforms should be liable for defamation on their sites. While the United States provides broad protections to platforms under Section 230 of the Communications Decency Act, which shields them from liability for third-party content, countries like India have started to demand much tighter rules that oblige platforms to more robustly seek out and remove defamatory content. The European Union is moving towards a more regulated approach, with obligations on platforms to remove illegal content swiftly once notified through its Digital Services Act. Thus, it presents a very complex legal challenge for governments of all countries across the world in the regulation of defamation on social media. A general aspect behind every country's legal framework is a particular balance between free speech and protection of reputation.

As such frameworks expand to accommodate an increasing amount of social media, it must evolve to address the unique issues raised by this online defamation area, such as liability between intermediaries, amplification of harmful content, and global spread. Comparative analysis of approaches to the law across countries, such as the United States, the United Kingdom, and India, show one diversity of responses, but there are also common issues in regulating defamation in the information society.

V. JUDICIAL PRECEDENTS

Courts around the world thus developed distinctive approaches to defamation actions that often reflect local approaches to procedure, common law or civil code approaches, and free speech protections. In the United States, First Amendment protections for freedom of speech form a

strong foundation for informing defamation decisions. The landmark case of New York Times Co. v. Sullivan set a high bar for proving defamation for public figures-to prove "actual malice": that the defendant knew the statement was false or acted in reckless disregard of the truth. In the context of social media, this precedent has further raised the bar even more substantially for public figures even when well-established harmful content spreads widely.

Defamation laws in the United Kingdom tend to be more protective of people's reputations than those in other jurisdictions. In fact, it was introduced in its 2013 Defamation Act[11] that a claimant would have to prove "serious harm" to reputation in order to bring action regarding defamatory content online. Some of the notable UK court rulings include making social media companies liable as well as individuals for making defamatory statements, an instance being a ruling where it was held that even the social media companies were liable for the defamatory statement when the writer, Jack Monroe, obtained damages against journalist Katie Hopkins due to the latter's tweets. In fact the courts proceeded to set a precedent by holding social media users liable for false and damaging statements shared online.[12]

Another very important legal development occurred within the ambit of social media defamation law in Australia. In **Fairfax Media Publications; Nationwide News Pty Ltd; Australian News Channel Pty Ltd v Voller[13]**, the High Court of Australia held that media houses could indeed be held liable for statements published on their respective social media pages under their accounts which could possibly amount to defamation. This ruling has far-reaching implications on content moderation and intermediary liability. The judgment imposed on the host an obligation to moderate comments on their platform notwithstanding such comments being posted by other persons than the owners of that respective platform.

The most important ruling at European jurisdiction is the judgment of the European Court of Human Rights, ECtHR, in its 2015 judgment popularly known as **Delfi v. Estonia[14]**. It was concerning the liability of a news portal online for defamation of remarks created by users who were anonymous. An Estonian news website, Delfi, published certain comments from the users, which included some defamatory and threatening statements. The ECtHR thus held Delfi liable for not deleting the posts in timely fashion as it was provided with marking software for inappropriate content. It was a judgment which codified, among other things, the terms of obligations of online service providers with regard to monitoring and disciplining user-

[11] Defamation Act 2013
[12] [2017] EMLR 16
[13] [2020] NSWCA 102
[14] Delfi AS v Estonia, App. No 64569/09 (ECHR 16 June 2015).

 [ISSN 2581-5369]

generated content, thereby changing, by an act of contractual sleight of hand, the connotation of intermediary liability across Europe.

This is defamation law in India. Though the framework may be a fruit of both the Indian Penal Code (IPC) and civil tort law. Under the provisions of Section 499, IPC, criminal defamation is a widely punishable offence and carries a sentence of two years. The Indian courts have had to fine-tune these laws to match the challenges presented by social media. As such, landmark cases such as **Subramanian Swamy v. Union of India**[15] held criminal defamation constitutional upon the ground that protecting an individual's reputation is a permissible curb on free speech. Indian courts have also addressed the issue of intermediary liability under the Information Technology (IT) Act. For instance, it was decided that, for instance in **Google India Pvt. Ltd. v. Visaka Industries**[16], the liability attached towards an intermediary such as social media sites for contents declared to be defamatory applied if they did not remove those from the website after notice.

These judicial precedents from all around the world, such as Delfi v. Estonia and India's defamation rulings, depict the changing legal landscape as courts are emerging with difficulties in the digital age of defamation. The issue of reconciling free speech with reputation on global platforms remains critical, and continuously evolving legal developments continue to shape the handling of defamation on social media.

VI. MONITORING RESULTS AND FUTURE RECOMMENDATIONS

Monitoring by social media about adherence to the Code of Conduct on Countering Illegal Hate Speech Online has been a mixed affair between progress and problems. The first round of monitoring, carried out by the European Commission in 2016, covered 12 civil society organizations working in nine Member States. Together, these 600 infractions of the code of conduct were reported by the three social media: Facebook, YouTube, and Twitter. The results were as follows; the average for removal of reported content was 28.2%. The actual figures from the three websites were: Facebook removed 28.3%, YouTube removed 48.5%, and Twitter removed only 19.1%. To boot, reports were addressed within the Code of Conduct's prescribed 24-hour window in 40% of cases.[17] A second scan in 2017 had proved much better.

Social media platforms removed 59% of the content reported, raising the proportion of complaints reviewed within 24 hours to 51%. By the end of the third auditing phase, completed

[15] Subramanian Swamy v. Union of India, (2016) 7 SCC 221
[16] Google India (P) Ltd. v. Visaka Industries, (2020) 4 SCC 162
[17] Jourová, Vera. "Code of Conduct on countering illegal hate speech online: First results on implementation." *Factsheet Directorate-General for Justice and Consumers* (2016).

at the end of 2017, the rate of removal was at 70%, and the response time had improved as well, with 81.7% of cases solved within 24 hours.[18] The outcome shows that although tremendous progress has been made regarding these social media sites, more should be done on timely removal and provision of feedback to users reporting violations.

Thus, reliance on user reports is another significant issue, as most platforms strongly rely on the users to flag hate speech. This does create an enforcement bottleneck since many of its users may not have the intention or knowledge of what content should be reported. In fact, results of monitoring show that YouTube and Twitter were more responsive to reports from "trusted flaggers" — those users with special expertise in detecting hate speech — than to reports from ordinary users, thus creating a gap in the effectiveness of content regulation.

Going forward, several recommendations will improve the regulation of hate speech on social media:. Firstly, user reports must be supplemented by active steps taken by the platforms themselves: automating systems to detect and thus flag hate speech. Secondly, platforms need to do better in transparency and accountability, especially in the form of returning meaningful feedback to the users reporting violations. This would steadily increase the communities' trust and enhance more users' participation in content moderation. Finally, invest in promoting counter-narratives to hate speech; the algorithms should be used not only for removing harmful content but also for amplifying messages that encourage inclusion, tolerance, and diversity.

General Trend of the Monitoring Exercised Social media platforms have got better at handling illegal content, and continued efforts and much stronger mechanisms are needed for comprehensive and equitable enforcement.

*****

[18] *Justice And Consumers,* Code of Conduct on Countering Online Hate Speech https://ec.europa.eu/newsroom/just/items/71674/en.

 [ISSN 2581-5369]