Applying Net neutrality rules to social media content moderation systems

Winston Maxwell

Applying Net neutrality rules to social media content moderation systems

By Winston MAXWELL
Director of Law and Digital Technology Studies,
Télécom Paris - Institut Polytechnique de Paris, Laboratoire i3 (UMR 9217)

I argue that the Net neutrality concept of "reasonable traffic management" can be applied to social media content moderation systems. Unlike recommendation systems which select, organize, and prioritize content, moderation systems should be neutral. Platforms should apply content moderation rules in an objective and non-discriminatory manner. The article explains the difference between content moderation and content recommendation (also called curation). The article then explores different forms of discrimination in content moderation. I propose two rules inspired by "reasonable traffic management" that should be transposed to content moderation: (i) discrimination in content moderation enforcement should not be motivated by commercial considerations, and (ii) discrimination should be based on objective criteria related to the nature of the content, the ease of detection and the relevant harms flowing from over-removal or under-removal. Finally, I argue that the proposed DSA should include the explicit requirements on the neutrality of content moderation, modeled on the language that appears on the European Regulation on the Dissemination of Terrorist Content Online.

The French Conseil d'État has proposed that social media platforms have a duty of fairness (*loyauté*) to users;[1] others have suggested a duty of neutrality.[2] The purpose of this article is to ask whether certain rules on Net neutrality can apply to the content moderation function of social networks. To answer this question, we first need to distinguish between the two functions of social media: their function as hosting provider, and their function as recommender of content. The hosting provider function is in theory passive. The platform accepts any user-generated content that is not prohibited by the platform's terms of use. When the E-Commerce Directive[3] was enacted, hosting providers allowed users to upload content without deploying tools to verify whether the content complied with the terms of use. Social media reacted to notices of harmful content *via* "notice and takedown" mechanisms. Content moderation has since become less passive. Hosting providers use machine-learning algorithms and large teams of human reviewers to analyze content

[1] French Council of State, Annual Study 2014, Fundamental rights in the Digital Age, (English summary), §II-2, 9 September 2014.

[2] Opinion no. 2014-2 of the French Digital Council on platform neutrality, Building an open and sustainable digital environment, May 2014.

[3] Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ('Directive on electronic commerce').

even before receiving a notice of a possible violation.[4] The content moderation function is designed to ensure that the content uploaded into the hosting space conforms to the terms of use, through either *ex ante* filtering or *ex post* review and removal.[5]

The recommendation function is quite different. Once content is in the hosting space, the recommendation function (also called content curation), prioritizes it to enhance each user's experience on the platform and thereby increase each user's engagement.[6] The recommendation system creates a personalized, sometimes addictive, user experience which generates profits for the platform, based on the selection and organization of content which, in theory at least, does not violate the terms of use. The recommendation algorithms are secret, and there is little neutrality in how recommendation systems operate.[7] They discriminate by design. Recommendation systems are the source of problems like addiction, filter bubbles, and manipulation of opinion, but neutrality is not the right regulatory remedy to address these problems.[8]

If we think of content moderation as a filter to keep harmful content out of the hosting space, the parallels with Net neutrality become evident. Internet access providers use reasonable traffic management measures to keep harmful traffic out of the network. The European Open Internet Regulation[9] prohibits Internet access providers from blocking or otherwise discriminating against traffic, unless the blocking or discrimination is necessary for "reasonable traffic management", *i.e.* measures to preserve the integrity and security of the network, of services provided *via* that network, and of the terminal equipment of end-users.[10] To be deemed reasonable, traffic management measures must be transparent, non-discriminatory and proportionate, and not based on commercial considerations. The proposed Digital Services Act (DSA) imposes similar conditions of transparency and objectivity on content moderation. Under the DSA, platforms would have to publish their rules on content moderation in clear and unambiguous language, and apply and enforce the rules in a diligent, objective, and proportionate manner, bearing in mind the users' fundamental rights under the EU Charter of Fundamental Rights. The European

[4] Content moderation processes are described in detail in Cambridge Consulting, The Use of AI in Online Content Moderation, 2019 Report Produced on Behalf of OFCOM, 18 July 2019.

[5] Art. 2(p), Proposal for a Regulation on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC,15 December 2020, COM(2020) 825 final (Proposed DSA), which defines content moderation as "activities undertaken by providers of intermediary services aimed at detecting, identifying and addressing illegal content or information incompatible with their terms and conditions, provided by recipients of the service, including measures taken that affect the availability, visibility and accessibility of that illegal content or that information, such as demotion, disabling of access to, or removal thereof, or the recipients' ability to provide that information, such as the termination or suspension of a recipient's account".

[6] Art. 2(o), Proposed DSA, defines recommender systems as a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service, including as a result of a search initiated by the recipient or otherwise determining the relative order or prominence of information displayed. On the difference between content moderation and content recommendation/curation, see E. Llansó, J. van Hoboken, and P. Leerssen, and J. Harambam Artificial Intelligence, Content Moderation, and Freedom of Expression, Transatlantic Working Group Paper, 26 February 2020.

[7] A. Candeub, Bargaining for Free Speech: Common Carriage, Network Neutrality, and Section 230, 22 Yale J.L. & Tech. 391, 430 (2020).

[8] J. Balkin, How to Regulate (and Not Regulate) Social Media, 1 J. of Free Speech L. 71 (2021).

[9] Regulation 2015/2120 of 25 November 2015 laying down measures concerning open internet access and amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services and Regulation (EU) No 531/2012 on roaming on public mobile communications networks within the Union (Open Internet Regulation).

[10] Art. 3(3), Open Internet Regulation.

Regulation on Dissemination of Terrorist Content Online[11] imposes similar conditions on content moderation systems used to detect terrorist content, requiring that platforms publish their terms and conditions prohibiting the dissemination of terrorist content, and apply the policies in a diligent, proportionate, and non-discriminatory manner with due regard to users' fundamental rights. The proposed DSA uses the words "diligent, objective, and proportionate", whereas the Regulation on Dissemination of Terrorist Content Online uses the words "diligent, proportionate, and non-discriminatory", but the intent is the same: content moderation policies should be articulated in clear terms and should be applied in a non-discriminatory manner with due regard to users' fundamental rights. As mentioned in the proposed DSA, content moderation should not yield "unfair or arbitrary outcomes".[12] The Regulation on Dissemination of Terrorist Content Online refers expressly to the need to take freedom of expression into account when applying content moderation mechanisms to terrorist content, in order to avoid over-removal.[13] As we will see in part four of this article, the Regulation on Dissemination of Terrorist Content Online also imposes more specific conditions than does the DSA with regard to the neutrality of content moderation.

The sections below will explore the parallel, which to my knowledge has not yet been explored, between reasonable traffic management measures in Net neutrality and non-discriminatory content moderation. The first section will present an overview of contact moderation systems, making a distinction between the terms of use that define the platform's rules on acceptable content, and the enforcement mechanisms used by social media to apply the rules. The second section will examine how discrimination can arise in content moderation systems. The third section will draw lessons from the Net neutrality concept of "reasonable traffic management". The fourth section will conclude, suggesting improvements to the proposed DSA.

Content moderation policies are private regulatory systems

The focus of this article is content moderation. Content moderation includes multiple elements, all of which constitute a private regulatory system.[14] The elements include the terms of use, internal guidelines to help human reviewers apply the terms of use, notice and takedown processes, algorithmic detection and filtering tools, teams of human reviewers, escalation procedures for complex cases, and complaint and appeal mechanisms.[15] This complex system can be divided into two main components: the set of rules defining what content is prohibited, and the mechanisms to enforce the rules. The terms of use are the platform's private laws defining prohibited content, and setting out the sanctions that might apply if prohibited content is uploaded in violation of the terms of use. The terms of use typically contain both broad standards, and precise rules. A standard is a flexible principle, such as prohibition of "offensive content", that lends itself to interpretation and can evolve over time. A rule is a precise provision, such as a prohibition of a photo of "uncovered female nipples", that requires little or no interpretation.[16]

[11] Regulation 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online.

[12] Recital 38, Proposed DSA.

[13] Art. 5(1), Regulation 2021/784 of 29 April 2021 on addressing the dissemination of terrorist content online.

[14] E. Douek, Content Moderation as Administration (January 10, 2022). forthcoming Harvard Law Review Vol. 136, Available at SSRN: https://ssrn.com/abstract=4005326

[15] See Cambridge Consulting, *supra* n. 4.

[16] On the distinction between rules and standards, see K. Clermont, Rules, Standards, and Such, 68 Buffalo L. Rev. 751 (2020).

A rule is easier for an algorithm to apply. A standard is more complicated, usually requiring human interpretation. Over time, the major social media platforms have made their terms of use more and more precise, responding to criticism that vague content standards leave too much room for discretion.[17] The Facebook community standards now describe prohibited content in detail. For example, instead of prohibiting images of "sexual acts" (a vague standard), the community standards provide a long list of examples of precise sexual acts (or simulations thereof) that are prohibited.[18] By contrast, Twitter's terms of use refer simply to "sexual acts"[19], leaving more room for interpretation. Being flexible, a standard won't require regular updates. A rule, on the other hand, needs regular updating to avoid becoming obsolete.[20] Smaller social media platforms still use broad standards, such as prohibiting "content that we consider to be offensive, objectionable, unlawful, explicit, graphic or otherwise in breach of our terms."[21] Terms of use generally prohibit both illegal content and content that is legal but violates the platform's policies. Illegal content is a smaller subset of a larger category of content prohibited by the terms of use.

The second component of content moderation consists of the enforcement mechanisms. Enforcement mechanisms consist of algorithmic detection tools and teams of human moderators.[22] Today, Meta says that 90% of prohibited content is detected by its algorithms, showing the heavy reliance on algorithms during the enforcement phase.[23] Algorithmic alerts can result in automatic blocking of content, or referral to human reviewers. Responding to criticism that human reviewers are not sensitive to local language, history, and culture, large social media platforms have deployed human moderation teams familiar with local conditions. The terms of use and filtering mechanisms may also differ depending on the region in which the user resides, permitting social media content moderation policies to adapt to local laws and culture. Complex content moderation questions may be escalated to a second team of reviewers. Users generally have the opportunity to challenge content moderation decisions. The proposed DSA would make the ability to challenge an absolute right.

The enforcement component of content moderation cannot be entirely separated from the recommendation system. One of the remedies applied by social media platforms for content they consider harmful if pushed to millions of people, is to downgrade the content in the recommendation system, a remedy that will limit the impact of the content without removing it entirely. This remedy is particularly relevant for misinformation campaigns, where the underlying content, *e.g.* a conspiracy theory, is a legitimate expression of an opinion, but its manner of propagation shows a deliberate and coordinated effort to manipulate opinions. As this example shows, there will be cases where recommendation and moderation overlap, which raises concerns of discrimination in moderation systems. Ideally, moderation should remain objective, unpolluted by the subjectivity of the

[17] European Commission Factsheet, Consumer Protection Cooperation Action on Facebook's Terms of Service, April 2019.

[18] Facebook Community Standards, Adult Nudity and Sexual Activity, https://transparency.fb.com/policies/community-standards/adult-nudity-sexual-activity/, visited on March 31, 2022.

[19] https://help.twitter.com/en/rules-and-policies/media-policy

[20] S. Breyer, Regulation and Its Reform, Harvard University Press, 1982.

[21] https://letterboxd.com/legal/community-policy/, visited on March 31, 2022.

[22] The use of algorithmic tools and human reviewers used by Meta for Facebook is presented here https://transparency.fb.com/enforcement/, visited on March 31, 2022. For a presentation of machine learning tools used for content moderation, see Cambridge Consulting, *supra*, n. 4, and R. Gorwa, R. Binns & C. Katzenbach, Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society. January 2020.

[23] Meta Transparency Center, How Technology Detects Violations, https://transparency.fb.com/enforcement/detecting-violations/technology-detects-violations/ visited on March 31, 2022.

recommendation systems. In reality, perfect separation may prove impossible. But thinking of content moderation as a separate function from recommendation, the former governed by objective, non-discriminatory enforcement rules, the latter free from such rules, will help content moderation gain in credibility and effectiveness.

Discrimination in content moderation

Content moderation should ideally be the non-discriminatory application of a set of rules designed to keep harmful content out of the hosting space. Yet, discrimination can still seep into the content moderation process, in three ways. First, discrimination can arise in the terms of use themselves; second, discrimination can occur in enforcement of the terms of use as a result of a deliberate decision by the platform managers; third, discrimination can occur in enforcement of the terms of use through unintentional bias.

Discrimination in the terms of use themselves is rare. It is possible in theory that a specialized social media site could, for example, restrict its service to members of a certain religion, and explicitly ban content that is offensive to that religion. I have not seen examples of this, or analyzed whether such a restriction would violate anti-discrimination laws and the EU Charter.[24] The terms of use of large social media prohibit content that is either illegal or harmful to a significant proportion of users, without singling out particular political or religious points of view. One form of deliberate discrimination in the terms of use may flow from regional differences: social media platforms may adopt different regional versions of their terms of use, reflecting local differences in law and culture. But overall, terms of use are neutral on their face.

Intentional (direct) discrimination in the enforcement of the terms of use results from a conscious decision to enforce a certain kind of violation, or sanction a certain person or group of persons, or on the contrary a decision not to enforce the terms against a certain person or group. An example of the former might be Meta's decision to suspend the account of former President Trump.[25] An example of the latter might be Meta's decision, reported by Reuters, to tolerate messages calling for violence against Russian soldiers.[26] Another deliberate form of discriminatory enforcement might be a delisting of an entity from search results because the entity represents an economic threat to the search engine operator.[27] These actions result from a deliberate enforcement (or non-enforcement) of the terms of use in a discriminatory way. The literature on content moderation is rife with examples of enforcement decisions that appear politically motivated, leading some scholars to argue that content moderation is subjective and discriminatory by nature, and can never be neutral.[28] The apparent subjectivity and political bias in enforcement decisions makes the moderation process inherently suspect. By clearly separating moderation from recommendation, and imposing an objectivity requirement on moderation, the proposed DSA attempts to remove, or at least reduce, the subjectivity problem.

Unintentional (indirect) discrimination is less discussed, but no less present in moderation processes. Discrimination can arise from algorithmic and human bias resulting in

[24] Such a discrimination on the basis of religion might be illegal under the CJEU's Egenberger decision, Case C-414/16 of 17 April 2018.

[25] Facebook Oversight Board decision May 5, 2021, Case decision 2021-001-FB-FBR.

[26] M. Vengatti & E. Culliford, Facebook allows war posts urging violence against Russian invaders, Reuters.com, March 11, 2022.

[27] This was the allegation made by the plaintiff in e-ventures Worldwide v. Google.

[28] A Chandler & V. Krishnamurthy, The Myth of Platform Neutrality, 2 Geo. L. Tech. Rev. 400 (2018); C. Castets-Renard, Algorithmic Content Moderation on Social Media in EU Law: Illusion of Perfect Enforcement, J. of Law, Technology & Policy 283 (2020).

certain kinds of content, or certain groups, being subject to stricter enforcement measures than other similar content or groups. For most social media, the policy prohibiting the posting of nude photos is applied strictly, in large part because it is easy for an algorithm to detect bare breasts or genitals. A policy prohibiting hate speech will generally be applied more loosely, not because hate speech creates less harm, but because fake news is harder to detect, requiring analysis of context and culture. The sensitivity level of a given content policy might also change over time. Near election periods, policies prohibiting political disinformation may become more strict. The decision of where to set the cursor for a given content policy, of how many human moderators to deploy, and how aggressively to use algorithmic tools, may be justified by objective arguments, for example a balancing of risks during election time, or during a pandemic.

Discrimination may arise when the same content policy is applied differently for different groups of the population depending (for example) on their language, nationality, ethnic origin, gender, religion, or political orientation. Evidence of unintentional algorithmic and human bias abound.[29] Insults to certain population groups may be detected and removed more effectively than insults to other population groups. Insults expressed in certain languages may be detected and removed more effectively than insults expressed in other languages. Finally, hateful content can be engineered to avoid detection algorithms, through use of words and images that fool the moderation system.

A major dilemma for operators of content detection algorithms is whether to prioritize over-removal errors (false positives) or under-removal errors (false negatives). For classification algorithms, there is generally a trade-off between the level of false positives and false negatives.[30] Operators of the algorithm cannot reduce both errors at the same time. They have to make a decision on which kind of error is worse in a given situation, and set the algorithm so that it strikes the right balance. A strict content removal policy will result in a high rate of false positives whereas a loose policy will result in a high rate of false negatives. A false positive harms the freedom of expression of the person whose content was wrongfully removed. A false negative harms the victims of the content that should have been removed but wasn't, for example the owner of a copyright, or the victim of a revenge porn photo. For some kinds of harmful content (*e.g.*, election manipulation), the victims may include society as a whole.

A common form of bias in image classification occurs when the system has a higher error rate when classifying images of dark-skinned people compared to light-skinned people, or *vice versa*. The same problem occurs in natural language processing, where classification error rates for certain languages will be much higher compared to others. Equalizing the error rate across groups or languages can result in a decrease in performance for all groups or languages.[31]

The CJEU and the French Constitutional Council have shown low tolerance for government-imposed measures that would result in over-blocking of content.[32] These cases only apply to measures directly imposed by the government. Laws that encourage platform

[29] R. Binns, M. Veale, M. Van Kleek, & N. Shadbolt, Like trainer, like bot? Inheritance of bias in algorithmic content moderation, arXiv:1707.01477, 2017.

[30] A. Tharwat, Classification assessment methods, Applied Computing and Informatics, Vol. 17 No. 1, 2021 pp. 168-192.

[31] S. Clémençon & W. Maxwell, Why facial recognition algorithms can't be perfectly fair, The Conversation, 20 July 2020.

[32] CJEU case C-70/10, Scarlet Extended v. SABAM, 24 November 2011; French Constitutional Council decision n°2020-801 DC of 18 June 2020 on the Law on fighting hate content on the internet.

operators to implement effective and proportionate measures[33] to limit the sharing of illegal content have so far not been invalidated on the ground that they lead to over-blocking. This is presumably because the platform operator is supposed to do its own analysis of proportionality before deploying the measures.[34] The over-blocking is not the direct result of a government order, but of private measures, like anti-spam filters, designed to protect social network users. The Regulation on Dissemination of Terrorist Content Online calls on platform operators to take due account of freedom of expression and to avoid over-blocking. Thus if over-blocking occurs, it is not the State's fault.

Lessons from Net neutrality

One of the lessons from Net neutrality is that traffic management measures cannot be based on commercial considerations.[35] If this rule were transposed to content moderation, it would mean that hosting platforms would not be able to enforce their content moderation policies differently depending on commercial considerations, such as whether the relevant content is likely to generate higher revenues for the platform. Commercial discrimination of this kind would be permitted within the recommendation system, but not at the level of content moderation. Ideally functional separation would divide the two roles to ensure that commercial strategy does not affect content enforcement decisions.[36] That is not to say that the likely impact and popularity of content could not be a factor in an enforcement decision: prohibited content with high impact and a high likelihood of going viral might justify quicker and stricter enforcement than similar content with low impact. However, this justification would be based on the likely harm resulting from the content, not on the commercial effect that removal (or non-removal) of the content would have on the platform's revenues. This rule would be critical to ensure that platforms cannot negotiate commercial deals in exchange for differentiated treatment by content moderation tools. (I have seen nothing to suggest that such deals exist, but they were a major concern in the Net neutrality debate.) It would also help ensure that commercial and ideological considerations, such as promoting a certain presidential candidate or promoting the social media group's own services, do not pollute content moderation decisions.

The second lesson from Net neutrality is that discriminatory treatment of traffic must be justified based on objective differences in technical service requirements for different categories of traffic. Transposed to content moderation, this would mean that discriminatory enforcement of content moderation policies should be justified by objective differences in:

- the nature of the content and the ease with which it can be identified with automatic tools;
- the likelihood of false positives and the harm associated with false positives for the relevant content;
- the likely harms associated with not removing the content (false negatives).

[33] French law of 24 August 2021, article 42 ("reasonable, effective and proportionate"); European Regulation on Dissemination Terrorist Content Online ("effective, targeted, proportionate"); Proposed DSA ("reasonable, proportionate and effective").

[34] W. Maxwell, The GDPR and Private Sector Measures to Detect Criminal Activity, Revue des Affaires Européennes - Law and European Affairs, Bruylant/Larcier (2021).

[35] CJEU judgment of 15 September 2020, Telenor Magyarország, C-807/18 and C-39/19, EU:C:2020:708, paragraph 48; CJEU judgments of 2 September 2021, Vodafone GmbH, C-854/19, C-5/20 and C-34/20.

[36] E. Douek, Content Moderation as Administration, *supra* n. 14.

As pointed out by the Regulation on the Dissemination of Terrorist Content Online and the proposed DSA, the consideration of harms should include harms to freedom of expression and other fundamental rights, such as non-discrimination. Consideration of the relative harms of different classes of prohibited content should be addressed anyway in the risk assessments conducted by very large platforms pursuant to Article 26 of the proposed DSA. The DSA's risk assessment would feed into the enforcement policy for content moderation, justifying differentiated enforcement policies based on objective factors.

Transposing the two Net neutrality rules relating to reasonable traffic management – *i.e.* discrimination in content moderation enforcement should not be motivated by commercial considerations, and should be based on objective criteria related to the nature of the content, the ease of detection, and the relevant harms flowing from over-removal or under-removal – seems consistent with the language of both the proposed DSA and the Regulation on the Dissemination of Terrorist Content Online. These regulations call for transparency, proportionality, objectivity (in the case of the proposed DSA), and non-discrimination (in the case of the Regulation on Terrorist Content). Being limited to content moderation, these rules of neutrality would not interfere with platforms' freedom to discriminate *via* their recommendation systems, including promoting certain content for commercial reasons. However, content moderation would be separated from commercial considerations, focusing only on the harms to users and to society flowing from removal, *versus* non-removal of the content, and the ease with which content can be detected and removed without excessive error. This would take a step toward functional separation of content moderation recommended by Douek.[37]

Improvements in the proposed DSA

The proposed DSA requires that removal decisions be accompanied by justifications, with reference to the specific provisions of the terms of use that were violated. The proposed DSA also imposes transparency obligations, requiring platforms to publish information on the content moderation algorithms they use, indicators of their accuracy and safeguards applied.[38] What's missing in the proposed DSA is a requirement that platforms test their content moderation systems for bias, both human and algorithmic, and implement steps to mitigate the identified biases. Intentional discrimination, such as suspending the account of one political party but not of another conveying similar extreme messages, would likely be contrary to the obligation of applying content moderation in a diligent, objective, and proportionate manner, respectful of freedom of expression. Discrimination based on a political point of view would presumably not be justifiable by objective differences in the harms to users and society of suspending one group's account *versus* another's. The selective suspension of the account could also be challenged as an unfair commercial practice.

Unintentional bias, such as unequal enforcement of content moderation policies based on the language used, will be more challenging. These biases will require more systemic measures, similar to measures that would be imposed on providers of high-risk AI systems under the proposed AI Act[39], including testing for bias, identifying biases, and developing mitigation measures. Surprisingly, content moderation algorithms, even for major platforms, escape most of the provisions of the proposed AI Act because they are not currently considered "high risk". If they remain outside the material scope of the AI Act,

[37] *Ibid*.

[38] Art. 23(1), Proposed DSA.

[39] Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, 21 April 2021, COM/2021/206 final (AI Act).

such algorithms should be subject to closer scrutiny under the proposed DSA, with a view to identifying and reducing biases.

The language of the proposed DSA on content moderation should be harmonized with the corresponding language in the Regulation on Dissemination of Terrorist Content Online. Each imposes neutrality-like obligations on content moderation systems. The Regulation on Dissemination of Terrorist Content Online is more specific on how a content moderation system should be applied in a "neutral" manner, requiring that systems be:

- effective in mitigating the level of exposure to the prohibited content;
- targeted and proportionate, taking into account the seriousness of the harms flowing from the content, and technical and operational capabilities;
- applied in a manner that takes full account of users' fundamental rights, including freedom of expression;
- applied in a diligent and non-discriminatory manner.[40]

The proposed DSA is much less specific, referring to objectivity and proportionality but regrettably avoiding the word "non-discriminatory". Yet content moderation systems, whether for terrorist content or other forms of harmful content, should apply the same standards of neutrality, basing enforcement decisions on objective criteria, such as those listed in the Regulation on Dissemination of Terrorist Content Online.

[40] Art. 5(3), Regulation on the Dissemination of Terrorist Content Online.