BRUSSELS PRIVACY HUB
Vrije Universiteit Brussel

The Digital Services Act, content moderation and elections

By Liubomir Nikiforov*

* PhD researcher in Law, Vrije Universiteit Brussel, lyubomir.nikiforov@vub.be

Contents

1 Digital Services Act

Some[1] have already labelled 2024 as a "high-risk" year due to the record-breaking number of upcoming elections worldwide. In Europe, the elections in several Member States and the EU Parliament Elections in June will call millions to the ballot boxes. This is going to be an exercise of democracy for the citizens and for the governments. However, it would be a challenge for the major social media platforms because the political ads flood would entail an additional pressure on the content moderation departments of those platforms. Social media platforms could be easily used with malicious intentions such as to influence social attitudes or incite hatred towards particular social groups or political opponents. Given the large impunity of hate crimes online, in some Europeans, countries, 23% of citizens have experienced cyber harassment in the last 5 years, with the EU-27 average being 14%[2]. Despite those statistics, tackling this issue is difficult due to the difficulty in monitoring, identification and elimination of this content. In order to answer to the growing concern around online hate speech, the EU introduced in 2016 the Code of Conduct on Countering Illegal Hate Speech Online[3], which requires major social media platforms to review and remove illegal content within 24 hours of notification. In spite of this soft law instrument, signed by some of the major social media platforms, the 24 hours assessing and deletion period of illegal material upon flagging has dropped from 90,4% in 2020 to 63,3% in 2022 according to FRA[4]. The poor results of the Code of Conduct evidenced the need of a different regulatory approach. Moreover, digital economy's evolution brought up by the development of new technologies, challenged existent EU digital legislation. This made the main regulatory framework for digital services, the e-Commerce Directive[5], adopted in 2000, insufficient.

---

[1] Shahi, M. (2023, September 14). *Protecting Democracy Online in 2024 and Beyond*. Americanprogress.Org. https://www.americanprogress.org/article/protecting-democracy-online-in-2024-and-beyond/

[2] This data is based on a FRA survey conducted in 2019, https://fra.europa.eu/en/data-and-maps/2021/frs, accessed on 6 January 2024

[3] European Commission. (2016, June 30). *Code of Conduct on Countering Illegal Hate Speech Online*. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

[4] *Online content moderation—Current challenges in detecting hate speech* (doi: 10.2811/923316; pp. 1–98). (2023). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/en/publication/2023/online-content-moderation#publication-tab-0

[5] 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'), OJ L 178 (2000). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32000L0031

In this context, the Digital Services Act (DSA)[6], already in force since 2022, is going to play an important role, due to its application start date on 17 February 2024. Although it does not replace, but upgrades, the current regulatory framework, the DSA is considered an important milestone in the fight against disinformation and hate speech[7]. An important remark is that the DSA does not define disinformation and hate speech as illegal content due to the diverse nature of those acts and their definition in other EU instruments or on a Member States level.

The Regulation primarily concerns online intermediary services (Art. 2). Art. 3 (g) provides a technical description of the term "intermediary service" as well as to online platforms (Art 3 (i)) and online search engines (Art. 3(j)). Those are exemplified as "online marketplaces, social networks, content-sharing platforms, app stores, and online travel and accommodation platforms"[8]. However, for the sake of readability, this blog post will referred to "platforms", "online platforms" or "social media" because those popular names designate the online spaces where disinformation and illegal content sharing most often happens.

The Regulation also establishes service providers' liability conditions (Arts. 4 to 6) for any uploaded illegal content, they have actual knowledge of. A crucial distinction should be struck between, from one side, very large online platforms (VLOPs) and very large online search engines (VLOSEs)[9], and, from the other, all other service providers. This differentiation has its critics[10] but more importantly, it has a decisive impact on the liability obligations concerning content moderation and its potential removal. Platforms are obliged assess and remove flagged content through notice and action mechanisms (Arts. 16 and 17). However, the burden of proactive content moderation is placed on the very large platform. Intermediary service providers are exempt from carrying out a proactive content monitoring, pursuant Art. 8. While

[6] Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277/1, 65 3 (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2022:277:FULL&from=EN

[7] van Hoboken, J., Quintais, J. P., Appelman, N., Fahy, R., Buri, I., & Straub, M. (Eds.). (2023). *Putting the DSA into Practice. Enforcement, Access to Justice, and Global Implications*. Verfassungsblog gGmbH. https://www.ivir.nl/publicaties/download/vHoboken-et-al_Putting-the-DSA-into-Practice.pdf

[8] European Commission. (2023, December 7). *The Digital Services Act package*. Digital-Strategy.Ec.Europa.Eu. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[9] According to Art. 33 DSA, VLOPs and VLOSEs should have an average number of users "equal to or higher than 45 million, and which are designated as very large online platforms or very large online search engines pursuant to paragraph 4" of the same Article. As of 18 January 2024, the list of the designated VLOPs and VLOSEs could be found here: European Commission. (2024, January 18). *Supervision of the designated very large online platforms and search engines under DSA*. https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses

[10] Erixon, F. (2021). "Too Big to Care" or "Too Big to Share": The Digital Services Act and the Consequences of Reforming Intermediary Liability Rules. *European Centre for International Political Economy*, *5*. https://ecipe.org/publications/digital-services-act-reforming-intermediary-liability-rules/

this means that platforms are not required to filter their servers for illegal content, they are encouraged to do so voluntarily in Art. 7. On the other hand, Art. 34 legally binds VLOPs and VLOSEs to "identify, analyse and assess any systemic risks" (Art. 34 (1)), issuing form their service. In this relation, very large service providers are obliged to act upon those risks by implementing the respective measures in order to reduce those risks, following Art. 35.

The DSA includes rules on users' fundamental rights protection (Arts. 1, 14, 34) as enshrined in the European Charter of Fundamental Rights (the Charter)[11]. In line with the Charter, the DSA explicitly refers to potential damages to democracy by disinformation and hate speech during an election period in Art. 34 (1)(c) by including " any actual or foreseeable negative effects on civic discourse and electoral processes, and public security" as a potential systemic risk, which need to be taken into consideration by designated very large online platforms or very large online search engines.

2 Content moderation and elections

Assessing and banning illegal material is tricky and this is why many of the infringing content is not removed[12]. The opposite is also true when it comes to content removed by being falsely identified as illegal. The quantity of content posted, its context and language diversity, moderators' own (lack of) knowledge and biases as well as the deficiencies stemming from incomplete or lacking applicable national laws contribute to this situation. Although, the DSA directly refers to potential damages to democracy, major social media do not refer to political opinions as a basis for online hate[13]. This is why in view of the 2024 elections; it would be insightful to see how the Digital Services Package legislative instruments, in particular the DSA, apply and how responsible parties tackle with the following concerns:

1. Internal policies and the human factor

While the DSA contributes to filling the gap between Member States in terms of regulation combating discrimination and hate speech, it is not clear how social media select, organize and

[11] Charter of Fundamental Rights of the European Union, OJ C 202/389, European Union, 389 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12016P/TXT

[12] *Online content moderation—Current challenges in detecting hate speech* (doi: 10.2811/923316; pp. 1–98). (2023). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/en/publication/2023/online-content-moderation#publication-tab-0; AIContentfy team. (2023, November 6). *The role of AI in content moderation and censorship*. Aicontentfy.Com. https://aicontentfy.com/en/blog/role-of-ai-in-content-moderation-and-censorship

[13] *Online content moderation—Current challenges in detecting hate speech* (doi: 10.2811/923316; pp. 1–98). (2023). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/en/publication/2023/online-content-moderation#publication-tab-0

oversee content moderators. There are significant differences in the number of personnel engaged in content moderation through major social platforms[14] as well as their work conditions, and therefore in the resources those platforms invest in tackling with infringing content. A practical issue platforms already face, which would aggravate with the application of the DSA, is the specific interpretation of a particular content, reported under Art. 16 (2)(a) as infringing, when it comes to minor languages and specific social as well as political contexts. Handling those cases would require investing more in human content moderators with specific language skills and knowledge. It is not self-evident that platform would make this investment, instead of relying on current personnel and automated moderation. Apart from the human resources issues, it is unclear what internal policies moderators follow and how they are applied. It is yet to be seen how the transparency mechanisms laid down in the DSA would work out.

Therefore, it must be ensured that platforms apply adequate internal policies and have sufficiently trained human resources to tackle disinformation and illegal content, in particular, during election campaigns.

2. AI moderation

Another concern is the use of Artificial Intelligence (AI) in content moderation. First, in order to be efficient, AI moderation systems have to be trained on large datasets, which are primarily available on specific languages, which automatically puts all users of less spoken languages in a disadvantaged position. Second, it is unclear how platforms use AI to tackle with infringing content and to what extend it is used in content moderation[15]. While AI may be helpful to deal with the large amount of content under review, it is far from being employed autonomously, also because AI moderation systems may be biased. In fact, it has been proven that AI demonstrates and perpetuates persistent biases and societal stereotypes based on specific characteristics such as cultural and foreign background, sexual orientation and religion towards particular groups in society[16]. Third, users, more often than not, are unaware of the AI biases and potential discrimination in place.

[14] Cite FRA report.
[15] Quote report OR BETTER ANOTHER RESOURCE
[16] *Bias in algorithms—Artificial intelligence and discrimination* (doi: 10.2811/25847; pp. 1–106). (2022). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

Thus, more efforts should be put to limit discrimination stemming from the training data used and the inherent biases contained in the AI system, especially, when it comes to elections and election candidates from minorities.

3. Liability

Faced with the challenge of content moderation, platforms are only held liable for the content on their service if they have actual knowledge of the illegal activity (Articles 5(1)(e) and 6(1)). Platforms are not obliged to perform any illegal content monitoring (Art. 8). However, they can automate this process. Hence, it is not clear how the automated content moderation transparency measures in Art 15 (e) would help in elucidating platform's knowledge of a particular illegal content, and therefore its liability. Thus, platforms may turn into a repository for hate speech and illegal content, which would otherwise be persecuted in VLOPs.

Therefore, more clarity on this question should enhance trust, accountability and efficiency in content moderation.

3 Conclusion

The upcoming European and global elections pose a critical challenge to major social media platforms regarding the potential misuse of these platforms for malicious intents like influencing social attitudes or inciting hatred. This situation is compounded by the difficulties in monitoring and eliminating such content as well as in ensuring a transparent and accountable political targeting and advertising.

The Digital Services Act, in force since 2022 and set to be fully applied on 17 February 2024, emerges as an important legislative framework in addressing disinformation and hate speech online. When it comes to political advertising, the European Commission has already proposed a Regulation on the transparency and targeting of political advertising[17] back in 2021. The Regulation complements the DSA when it comes to ensuring a transparent and accountable political targeting and advertising and introduces additional obligations to political ads. publishers as well as service providers. Although it was initially meant to be applicable from 1 April 2023, it is still under EU Parliament's review. This underlines the crucial role of the DSA during this year's elections season.

[17] Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising COM/2021/731 final, (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0731

Challenges persist, however, in the implementation of the DSA, particularly concerning content moderation during elections such as the applicable internal policies, the human factor, the use of AI systems and the determination of platforms' liability. Addressing these concerns demands a multi-faceted approach involving robust internal policies, investments in diverse human resources, careful oversight of AI systems, and clearer delineation of platform liability. The effective implementation of the DSA, in tandem with these measures, will be paramount in safeguarding democratic processes and mitigating the spread of harmful content during crucial election periods.

4 References

European Commission. (2024, January 18). *Supervision of the designated very large online platforms and search engines under DSA*. https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses

European Commission. (2023, December 7). *The Digital Services Act package*. Digital-Strategy.Ec.Europa.Eu. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

AIContentfy team. (2023, November 6). *The role of AI in content moderation and censorship*. Aicontentfy.Com. https://aicontentfy.com/en/blog/role-of-ai-in-content-moderation-and-censorship

Shahi, M. (2023, September 14). *Protecting Democracy Online in 2024 and Beyond*. Americanprogress.Org. https://www.americanprogress.org/article/protecting-democracy-online-in-2024-and-beyond/

European Commission. (2016, June 30). *Code of Conduct on Countering Illegal Hate Speech Online*. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en

Online content moderation—Current challenges in detecting hate speech (doi: 10.2811/923316; pp. 1–98). (2023). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/en/publication/2023/online-content-moderation#publication-tab-0

van Hoboken, J., Quintais, J. P., Appelman, N., Fahy, R., Buri, I., & Straub, M. (Eds.). (2023). *Putting the DSA into Practice. Enforcement, Access to Justice, and Global Implications*. Verfassungsblog gGmbH. https://www.ivir.nl/publicaties/download/vHoboken-et-al_Putting-the-DSA-into-Practice.pdf

Bias in algorithms—Artificial intelligence and discrimination (doi: 10.2811/25847; pp. 1–106). (2022). FRA – European Union Agency for Fundamental Rights. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf

Regulation (EU) 2022/2065 of the European Parliament and the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277/1, 65 3 (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L:2022:277:FULL&from=EN

Erixon, F. (2021). "Too Big to Care" or "Too Big to Share": The Digital Services Act and the Consequences of Reforming Intermediary Liability Rules. *European Centre for International Political Economy*, *5*. https://ecipe.org/publications/digital-services-act-reforming-intermediary-liability-rules/

Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising COM/2021/731 final, (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0731

Charter of Fundamental Rights of the European Union, OJ C 202/389, European Union, 389 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12016P/TXT

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'), OJ L 178 (2000). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32000L0031