



Commentary

Self-regulation 2:0? A critical reflection of the European fight against disinformation

In presenting the European Democracy Action Plan (EDAP) in 2020, the European Commission pledged to build more resilient democracies across the EU. As part of this plan, the Commission announced intensified measures to combat disinformation, both through the incoming Digital Services Act (DSA) and specific measures to address sponsored content online. Ostensibly, these reforms would end the era of haphazard self-regulation that has characterized the EU response to disinformation. However, purported changes in this area are vaguely framed, and fail to address critical issues such as the regulation of harmful but lawful content. While instruments like the DSA show signs of improvement, shortcomings in this evolving framework represent a continuation of the EU's piecemeal approach to disinformation.

Author: Ethan Shattock

Affiliation: Department of Law, Maynooth University, Ireland

How to cite: Shattock, E. (2021). Self-regulation 2:0? A critical reflection of the European fight against disinformation. *Harvard Kennedy School (HKS) Misinformation Review*, 2(3).

Received: March 4th, 2021. Accepted: May 15th, 2021. Published: May 31st, 2021.

Decoding the Codes of Practice (CPD)

Since 2018, the European Commission has driven the legislative agenda for online disinformation through the Code of Practice on Disinformation (CPD). In a nutshell, technological signatories including Facebook, Twitter, and more recently TikTok, voluntarily commit to minimizing disinformation and coordinated election interference. Platforms agree to secure services against inauthentic behavior, encourage transparent “issue-based advertising,” and share relevant data to the “research community” (European Commission, 2020).

As a self-regulatory framework, platforms are not obliged to implement any specific practice but agree to report their activities to the Commission. Ultimately, incentives for implementing the CPD are largely predicated on reputation and evading regulation (Bertolini et al., 2021). Accordingly, there is no binding framework “specifically designed to tackle disinformation online,” (De Cock Buning et al., 2018). The CPD review system audits five key areas. These are the scrutiny of advertising, issue-based advertising, integrity of services, empowering consumers, and empowering the research community. The monitoring process itself is somewhat fractured. Numerous bodies are responsible for assessing the CPD. The European Regulators Group for Audiovisual Media Services (ERGA), the Commission, the signatories, and

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

third-party consultancies all share the burden of measuring the CPD's effectiveness. Moreover, the assessment criteria have drifted since 2018. Initially, the focus was on the integrity of the 2019 European Parliamentary elections, while more recently the focus has shifted to COVID-19.

Two years into the CPD, results have been mixed. While the Commission has commended “comprehensive efforts” by signatories to engage fact checkers, it has criticized the varied “speed and scope” in how the CPD have been implemented across platforms. All platforms require political advertisements to display a “paid for by” sign. However, signatories vary widely in their definitional scope of “political.” This variance has led to a lack of “consistent implementation of specific restrictions” for political advertising (European Commission, 2020). It is also arguable that the CPD's focus on political advertisements could detract from other vectors for European disinformation, as many forms of manipulated content do not stem from paid political content (Marwick & Lewis, 2017; Basch et al., 2021).

While an important pillar of the CPD is to “empower the research community” in identifying trends related to disinformation in the EU, the current provision of data falls short of meeting “the needs of researchers for independent scrutiny.” Platforms that *have* developed “repositories of political ads” retain the ability to unilaterally “alter or restrict” repository access. This has led the Commission to criticize the “episodic and arbitrary” access to researchers which in turn has obscured the “searchability” of relevant data, fostering knowledge gaps that prevent researchers from ascertaining “persistent or egregious purveyors of disinformation” in Europe (European Commission, 2020). Even if this data was more accessible, a limitation is that the reporting on “inauthentic” behavior is collated at a global level, making it difficult to understand specific disinformation campaigns relevant to EU Member States.

An overarching problem is the intrinsic limitation of self-regulation. The voluntary nature of the CPD does not promote a concrete “structured cooperation between platforms” (European Commission, 2020). Platforms do not face material sanctions for implementation failures. The most severe consequence is potential expulsion from the Codes, and accompanying reputational damage. The criteria for assessment, often carried out by the signatories themselves, insufficiently addresses the protection of fundamental rights. The targeted and efficient spread of disinformation in Europe is arguably a form of electoral interference that undermines the right to free elections, particularly in light of well-documented evidence that anti-democratic actors have targeted voters with extremist and xenophobic content in the run-up to European parliamentary elections and elections at a domestic level (Pierri & Ceri, 2021; Ferrara, 2017). The need for free elections and non-discrimination are codified into legally binding EU human rights instruments such as the Charter of Fundamental Rights (CFREU), and have been implemented by influential courts of review such as the Courts of Justice of the European Union in Luxembourg and the European Court of Human Rights in Strasbourg. Accordingly, it is highly questionable whether the protection of rights from disinformation should be reserved to the purview of private platforms in lieu of permanent regulatory oversight.

Disinformation in the Digital Services Act (DSA): Light touch liability?

The CPD sits within a broader evolving framework in the EU. Since 2000, the flagship instrument for regulating digital services in the single market has been the Electronic Commerce Directive. Arguably, the defining feature of the Directive is not its obligations, but its liability exemptions. Articles 12-15 exempt providers from liability for third-party unlawful content, conditional in that those providers must act “expeditiously” to remove such content. Once providers take steps to remove illegal content, they are effectively absolved from secondary liability for illegal content disseminated by users.

The consensus now is that this Directive is obsolete. Much has changed in Europe’s communication landscape since 2000, and new platforms yield unprecedented opportunities and threats (Borgesius et al., 2018). Unsurprisingly, these shifts have prompted calls for reform in order to align legislation with social

media (De Streel, 2018). In response to these calls, the Commission pledged to modernize platform regulation in its “Shaping Europe's Digital Future” commitment. In 2020, the Commission unveiled the long-awaited Digital Services Act (DSA). The DSA, in a nutshell, is an attempt to recalibrate intermediary responsibilities for curbing illegal content, in light of contemporary technologies.

The DSA differentiates between different classifications of service providers. These are broken down into:

- Intermediary services (internet access providers, domain name registrars)
- Hosts (cloud and web hosting services)
- Online platforms (app stores and social media platforms)
- Very large online platforms (platforms reaching more than 10% of 450 million monthly European consumers)

Certain provisions of the DSA apply to all providers. For example, *all* providers are subjected to new transparency obligations, and *all* providers must establish a “single point of contact” (SPoC) in the EU. However, certain obligations are tailored according to the classification of providers. “Hosts” must furnish transparent rules related to notice and takedown mechanisms, and disclose reasons underlying decisions to disable access to illegal content. This information must be made available in a database controlled by the Commission. Online platforms and very large online platforms submit to a “trusted flagger” regime, where flaggers notify platforms of illegal content and platforms can act on this “with priority and without delay.” This formalizes Notice and Takedown (NTD) procedures into binding EU law. NTD procedures have been inculcated into domestic law in Europe, for example in Germany’s Network Enforcement Act (NetzDG). The DSA necessitates that large platforms establish trusted flaggers to initiate this process as part of an internal complaints system. Larger platforms also have specific obligations to verify the identity of advertisers and disclose relevant information related to profiling in advertising procedures. This is a notable development in light of the role of tailored advertisements in the Cambridge Analytica scandal (Cadwalladr, 2018).

The DSA has been labeled as the EU’s “most ambitious plan yet to rein in online platforms” (Milo & Kreko, 2021). However, an appraisal of the DSA as a watershed moment for European intermediary liability could be premature. There *are* undeniably positive signs in the DSA pertaining to disinformation. There is a shift of focus from the narrow category of “political advertisements” to the broader scope of “paid for” content. There is also scope for the implementation of the CPD to be assessed as part of systemic risk management, which could carve room for robust oversight in how anti-disinformation measures are implemented. This has led to the DSA being characterized as a “co-regulatory backstop” for disinformation (Tambini, 2021). However, important aspects of platform responsibilities remain unclear. The cornerstone of liability exemption remains fundamentally unchanged in the DSA, as the regulation does not envisage general monitoring obligations to “actively . . . seek facts or circumstances indicating illegal activity.” In its currently proposed format, the DSA introduces transparency requirements for political advertising. While language has shifted, a major misconception is the interchangeable association between political ads and disinformation. Political advertising is only a fraction of this problem. European disinformation campaigns often spring from organic users, and counterfeit news sites (Bennett & Livingston, 2018). Attempts to focus on countering disinformation through political advertising will also run into the problem that domestic legislation for political advertisements lacks uniformity across EU Member states. As highlighted by the European Court of Human Rights (ECtHR) in *Animal Defenders v United Kingdom*, there is “no European consensus between the contracting states on how to regulate paid political advertising” (Kleinlein, 2017). Furthermore, an over-emphasis on political advertising is in itself problematic, as this represents only a subset of disinformation as a wider problem, not the problem in its entirety.

The quagmire of harmful but lawful content

A problem unlikely to be resolved by the DSA is the regulation of harmful but lawful content. Unlike child pornography and copyright infringement, disinformation is often not illegal *per se*. While other online harms are subjected to binding rules in Europe, disinformation is relegated to piecemeal soft law. A pervasive concern often raised when considering harmful but lawful content is that regulating such content could undermine freedom of expression, and this argument has cropped up when positing disinformation regulation (Smith, 2019).

These concerns are somewhat misplaced. If disinformation is *not* addressed with recursive legislation, the governance of disinformation will be reserved for commercial platforms. While the removal of unlawful content entails clear-cut responsibilities, the lack of concrete obligations to remove disinformation leaves wide discretion. Accordingly, a chief concern is that "notice and takedown" regimes allow platforms to indirectly regulate expression. This was crystallized with the NetzDG in Germany. Under this law, platforms must remove unlawful content within a 24-hour period, at the risk of financial sanctions. This elicited criticism that characterized the NetzDG as a "vague" and "overbroad" mechanism that "turns private companies into overzealous censors to avoid steep fines, leaving users with no judicial oversight or right to appeal" (Human Rights Watch, 2018). In this connection, an absence of legal oversight on account of the *lawful* nature of disinformation will not necessarily safeguard fundamental rights, and it is conceivable that a continued regulatory vacuum could further exacerbate freedom of expression concerns.

The gap for harmful but lawful content is evidenced in the DSA. The trusted flaggers notify platforms of content that is unlawful. Reporting mechanisms for notice and takedown procedures are directed towards unlawful content. While the risk assessment obligations for very large online platforms *are not strictly limited to unlawful content*, the scope of systemic risks is widely defined, emboldening anxieties surrounding how content moderation decisions could adversely affect digital speech (Kuczerawy, 2019). Many domestically sourced websites and blogs will not be subjected to these risk assessments, despite the reality that smaller platforms remain influential in diffusing disinformation in European Parliamentary elections (Pierri & Ceri, 2021). The penetration of domestic websites through organic user interaction is an important way in which "radical right" parties have capitalized on waning "institutional legitimacy" to influence elections such as the Brexit referendum (Bennett & Livingstone, 2018). Even the European Commission has acknowledged the DSA's limitations in this respect, noting that the regulation "will not explicitly address some of the very specific challenges related to disinformation." Instead, more tailored changes are reserved to the updated Codes of Practice on Disinformation (European Commission, 2020).

The current European legal framework for disinformation remains linked to a model that has already proved ineffective, that of self-regulation (Madiaga, 2020). This reality exists in spite of the illusory reform under the DSA. While tempting to posit direct regulation, this prospect is fraught with legal, political, and territorial challenges. The EU is a collection of 27 Member States, each possessing unique political environments wherein different country-specific disinformation campaigns diffuse (Bayer, 2019). An aggressive top-down attempt to harmonize strict rules for harmful but lawful content could destabilize political cohesion in the Union. In addition, there is a need for further debate on how to reconcile the regulation of harmful but legal content with fundamental rights to freedom of expression.

What can be done in the interim is to smoothen the transition from self-regulation to co-regulation in a manner that rectifies gaps in the EU fight against disinformation. There needs to be a reassessment as to whether disinformation can realistically be subsumed into the DSA in its current format. At present, the Commission has committed to revamping the CPD in 2021. However, the Commission has simultaneously recognized that many of the CPD's failures stem from their self-regulatory nature (European Commission, 2020). Therefore, it would be preferable to tackle disinformation through the

attempted co-regulation in the DSA. If it is deemed too late to reshape the DSA in a manner that puts more focus on disinformation, other instruments should be considered. The focus should be on firstly rectifying inconsistencies in the CPD. Consistent definitions and approaches need to be encouraged, and sanctions for non-compliance need to be considered. Oversight should scrutinize how co-regulation can improve anti-disinformation efforts while safeguarding fundamental rights. A natural starting point is that the scope of this problem, and its effect in the EU, need to be accurately understood and transparently communicated. Accordingly, the obscured access to relevant data for researchers should be lifted, respecting the contours of the General Data Protection Regulation (GDPR). Platforms should not use such instruments as a superficial rationale to avoid compliance with requests for data that can help ascertain their amplification of disinformation.

Conclusion

For disinformation in Europe, the era of self-regulation *appears* to be nearing a close. The CPD, while establishing important commitments, lacks enforcement and gives too much discretion to digital platforms to implement practices (Colliver, 2020). This discretion, as recognized by the European Commission, has led to severe implementation gaps that have prevented a coordinated response to online disinformation across Europe (Madiaga, 2020). As the announcement of the incoming Digital Services Act (DSA) would seem to suggest, there is a need to inculcate greater transparency and due diligence in the way that digital actors are accountable for harmful content on their platforms, and the European Commission has responded by bringing EU rules in line with the realities of contemporary digital engagement. However, with respect to *disinformation*, the devil is in the detail. Important gaps still continue to plague the response to disinformation, and the DSA fails to address many of these gaps.

A key question that should shift debates forward is whether harmful but lawful content should continue to eschew regulation purely because it is not illegal per se. As the European Union's "Democracy Action Plan" points out, the right to free and fair elections and the strengthening of "media freedom" should be front and center of the evolving agenda in this area. Accordingly, a number of important points must be addressed going forward. Firstly, while not all forms of disinformation are unlawful, some are. Some aspects of discriminatory and racist disinformation may well run contrary to both domestic legislation and EU law. Secondly, the imposition of binding rules for disinformation does not necessarily have to manifest through mandates for content removal. Tangible and concrete sanctions for platforms that fail to minimize disinformation could be a reasonable backstop. Obligations do not necessarily have to involve concrete take down mechanisms, and could focus on systemic approaches to minimizing disinformation, disincentivizing inauthentic behavior, and collaborating with researchers to identify persistent disinformation campaigns at an EU and EU Member State level. This could be achieved with the introduction of a new Directive that independently addresses grey areas and systemic risks associated with disinformation while still maintaining room for discretion as to how Member States implement harmonized rules at a domestic legislative level. Crucially, it must be pointed out that, under European human rights law, scrutiny of legal interferences with free speech is not merely focused on whether content was lawful or unlawful. As evidenced in a wide array of case law, both in EU and non-EU-related courts, legal interferences with harmful content are often centered on the legitimate democratic aim of restrictions, the legal precision of sanctions, and the proportionality of actions. Just because content may not be strictly illegal does not preclude any degree of binding regulatory scrutiny. This recognition has already been expressed by the European Commission, as current proposals in the DSA provide a scope for oversight of not only unlawful content but also content that poses risks to "public interests" and "fundamental rights" (European Commission, 2020).

At a bare minimum, there is a clear need for stronger oversight, harmonized approaches, and greater access to important knowledge on how and where disinformation crops up in the EU. Because of continuing shortcomings in the current framework, longstanding questions surrounding the trade-off between disinformation legislation and freedom of expression will, for now, continue to go unanswered (Helm & Nasu, 2021; Posetti & Bontcheva, 2020). Increasingly, it is clear that disinformation, in being a unique legal problem in Europe, requires a designated legislative agenda that finally graduates beyond self-regulation. Attempts to shoehorn reform into the DSA should not be mistaken for concrete progress.

Bibliography

- Basch, C. E., Basch, C. H., Hillyer, G. C., Meleo-Erwin, Z. C., & Zagnit, E. A. (2021). YouTube videos and informed decision-making about COVID-19 vaccination: Successive sampling study. *JMIR Public Health and Surveillance*. <https://doi.org/10.2196/preprints.28352>
- Bayer, J. (2019). *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States*. European Parliament Think Tank. [https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU\(2019\)608864](https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2019)608864)
- Bertolini, A., Episcopo, F., & Cherciu, N. A. (2021). *Liability of online platforms*. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU\(2021\)656318_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2021/656318/EPRS_STU(2021)656318_EN.pdf)
- Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139. [10.1177/0267323118760317](https://doi.org/10.1177/0267323118760317).
- Cadwalladr, C. (2018). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*. <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump>
- Colliver, C. (2020). *Cracking the code: An evaluation of the EU code of practice on disinformation*. Institute for Strategic Dialogue. https://www.isdglobal.org/wp-content/uploads/2020/06/isd_Cracking-the-Code.pdf
- De Cock Buning, M. (2018). *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation*. European Commission. <https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en>
- De Streel, A. (2018). Online intermediation platforms and fairness: An assessment of the recent commission proposal. *SSRN*. <http://dx.doi.org/10.2139/ssrn.3248723>
- Human Rights Watch. (2018). *Germany: Flawed social media law*. <https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law>
- European Commission. (2020). *Assessment of the code of practice on disinformation – Achievements and areas for further improvement*. <https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>
- Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *ArXiv*. <https://arxiv.org/pdf/1707.00086>
- Helm, R., & Nasu, H. (2021). Regulatory responses to 'fake news' and freedom of expression: Normative and empirical evaluation. *Human Rights Law Review*, 21(2), 302–328. <https://doi.org/10.1093/hrlr/ngaa060>

-
- Kleinlein, T. (2017). Consensus and contestability: The ECtHR and the combined potential of European consensus and procedural rationality control. *European Journal of International Law*, 28(3), 871–893. <https://doi.org/10.1093/ejil/chx055>
- Kuczerawy, A. (2019). *Fighting online disinformation: Did the EU Code of Practice forget about freedom of expression?* SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3453732
- Madiega, T. (2020). *Reform of the EU liability regime for online intermediaries: Background on the forthcoming digital services act*. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA\(2020\)649404_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2020/649404/EPRS_IDA(2020)649404_EN.pdf)
- Marwick, A., & Lewis, R. (2017). *Media manipulation and disinformation online*. Data & Society. <https://datasociety.net/library/media-manipulation-and-disinfo-online/>
- Milo, D., & Kreko, P. (2021). *Is the Digital Services Act a watershed moment in Europe's battle against toxic online content?* New Europe. <https://www.neweurope.eu/article/is-the-digital-services-act-a-watershed-moment-in-europes-battle-against-toxic-online-content/>
- Pierri, F., Artoni, A. & Ceri, F. (2021). Investigating Italian disinformation spreading on Twitter in the context of 2019 European elections. *PLOS ONE*, 15(1). <https://doi.org/10.1371/journal.pone.0227821>
- Posetti J., & Bontcheva K. (2020). *Disinfodemic: Deciphering COVID-19 disinformation*. UNESCO Policy Brief #2. <https://en.unesco.org/covid19/disinfodemic>
- Smith, R. (2019). Fake news, French law and democratic legitimacy: Lessons for the United Kingdom? *Journal of Media Law*, 11(1), 52–81. <https://doi.org/10.1080/17577632.2019.1679424>
- Tambini, D. (2021). *Media policy in 2021: As the EU takes on the tech giants, will the UK?* London School of Economics. <https://blogs.lse.ac.uk/medialse/2021/01/12/media-policy-in-2021-as-the-eu-takes-on-the-tech-giants-will-the-uk/>
- Zuiderveen Borgesius, F. J., Möller, J., Kruikemeier, S., Ó Fathaigh, R., Irion, K., Dobber, T., Bodo, B. & de Vreese, C. (2018). Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1), 82–96. <https://doi.org/10.18352/ulr.420>

Acknowledgements

The author would like to thank the editorial committee for reviewing this commentary. A special thanks is given to Costanza Sciubba Caniglia for providing helpful feedback on earlier drafts of this commentary.

Funding

Funding supplied by Maynooth University.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Competing interests

The author has no potential conflicts of interest.