CELE
Centro de Estudios en Libertad de Expresión

Documento de posición

34

CELE's input on recurrent and prominent systemic risks in the EU and on measures for their mitigation

Nicolás Zara

April 2025

Nicolás Zara, "CELE's input on recurrent and prominent systemic risks in the EU and on measures for their mitigation", Documento de posición No. 34 (ENG), Centro de Estudios en Libertad de Expresión (CELE), Buenos Aires (2025).

Facultad de Derecho

Centro de Estudios en Libertad de Expresión

UP
Universidad de Palermo

CELE's input on recurrent and prominent systemic risks in the EU and on measures for their mitigation

Nicolás Zara

CELE/UP

cele@palermo.edu

April 2025

Introduction

The Center for Studies on Freedom of Expression (CELE) is an academic research center affiliated with Universidad de Palermo in Argentina. The Center provides technical and legal analysis on issues affecting this fundamental right and has been studying freedom of expression on the Internet as a specific research area since 2012. The Center is a leading voice in promoting and protecting freedom of expression nationally, regionally, and internationally.

CELE has closely followed the drafting process of the Digital Services Act from its outset and, since 2024, has a dedicated research area focused on its study. Moreover, CELE is a member of the DSA Civil Society Coordination Group, led by the Center of Democracy and Technology, and of the DSA International Human Rights Alliance, led by the Electronic Frontier Foundation and AccessNow.

This submission was prepared in response to the invitation by the DSA team at DG-CNECT to provide input that helps obtain an additional perspective on the systemic risks and mitigation measures referred to in Articles 34 and 35 DSA, to inform the work on the report referred to in Article 35(2) DSA, to be issued by the European Board of Digital Services in cooperation with the European Commission.

We welcome the opportunity to contribute with an analysis that emphasizes freedom of expression. As an academic institution, we are grateful for the opportunity to participate

in this process and wish to underscore the vital role of academic organizations in such consultation efforts.

1. QUESTION 1. The report to be published once a year by the European Board for Digital Services in cooperation with the Commission pursuant to Article 35(2) DSA should outline the most recurring and prominent risks stemming from VLOPs and VLOPSEs.

1.A. Please provide any information you have that is suitable for identifying and assessing systemic risks you find potentially prominent or recurrent. The submission can consistent e.g. of studies (conducted by yourself or third parties), samples of typical constellations occurring at the use of the service and relevant findings or conclusions in regards of (typical) practical experiences made by users you represent or are aware of.

1.B. Where available, please include information about what makes the risk prominent or recurrent.

1.C. Please specify whether the information you provide relates to a single Member State, to several Member States or whether it applies to the entire Union.

1.D. Please refer to any existing documentation, research or resources that could help substantiate the evidence you provide.

Response to question 1:

Article 34 of the Digital Services Act requires VLOPs and VLOSEs to identify, analyze, and assess any systemic risks in the Union stemming from the design or functioning of their service and its related systems, including algorithmic systems, or from the use made of their services[1]. While Article 34.1 provides a set of systemic risks that must be mandatorily assessed and mitigated, it is not a closed number. Other systemic risks, different than those mentioned in Article 34.1 (a) to (d), that the legislator could not foresee, could exist or appear; the law entrusts VLOPs/VLOSEs with the task of identifying, assessing and mitigating these as well, for they are in a better position than state actors to do it. This flexible stance is part of the managerial co-regulatory approach of the DSA, which stems, in part, from the assumption that there is a significant knowledge imbalance between the regulator and the regulated subjects in favor of the latter[2].

[1] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and Amending Directive 2000/31/EC (Digital Services Act), article 34.1.

[2] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez-Ugarte, Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation (February 28, 2025). CELE Research Paper No. 64 (Forthcoming in the Journal of Intellectual Property, Information Technology and Electronic Commerce, vol. 16, No. 2 (2025), Available at SSRN: https://ssrn.com/abstract=5161173 or http://dx.doi.org/10.2139/ssrn.5161173

However, unlike the European AI Act[3], the DSA does not define *risk* nor indicate what makes one such risk *systemic*. While taking a closer look at the financial services literature –the field where the term was borrowed from– could be of use, the concept is not automatically extrapolable to platform regulation, as shown by Micova and Calef.[4] Moreover, a recent research piece by AlgorithmWatch showed that "researchers often reached very different conclusions about whether a given case was evidence of 'systemic risk', and how clearly this decision could be reached".[5] Definitions are desperately needed because the category remains "remarkably broad"[6] and practically unworkable. As shown by the first round of risk assessment reports, platforms have refrained from providing their definitions, which are unlikely to come up organically from the DSA iterative process. Moreover, risk is a polysemic concept as "there are deep ideological and political conflicts over the nature of these essentially contested concepts".[7] The European Commission, as the entity with legislative initiative within the European Union, must step up and start an extensive consultation process to advance legislation that defines key terms (what is a risk? what makes one such risk systemic?) and lays down objective parameters that help VLOPs/VLOSEs identify one when they see it. Not doing so will leave the DSA regime up for grabs for regulatory capture by platforms and would grant regulators an enormous amount of discretion, which could be used in the future to shape and control public discourse.

Some of the mitigation measures included in Article 35 of the DSA directly or indirectly affect the way human expression is distributed, organized, and treated (left up, removed, geoblocked, demonetized, viralized, and so on) by platforms. The obligation to identify and mitigate risks—vaguely defined and enforced through the threat of hefty penalties—is likely to drive platforms, as self-interested for-profit actors, to overcompensate by flagging and addressing anything that might remotely be considered a

[3] Article 3 (65):"'systemic risk' means a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain"

[4] Sally Broughton Micova and Andrea Calef, 'Elements for Efective Systemic Risk Assessment Under the DSA' [2023] SSRN Electronic Journal 50

[5] Oliver Marsh, 'Researching Systemic Risks Under the Digital Services Act' (AlgorithmWatch, 26 July 2024), p. 5

[6] Paddy Leerssen, 'Outside the Black Box: From Algorithmic Transparency to Platform Observability in the Digital ServicesAct' (2024), 4 Weizenbaum Journal of the Digital Society 24.

[7] Rachel Grifn, 'What Do We Talk about When We Talk about Risk? Risk Politics in the EU's Digital Services Act - DSA Observatory' (DSA Observatory, 31 July 2024) <https://dsa-observatory.eu/2024/07/31/what-do-we-talk-about-whenwe-talk-about-risk-risk-politics-in-the-eus-digital-services-act/>

risk. This is highly problematic if what is deemed "risky" is human expression, and that is why any action of state-mandated risk mitigation by platforms must survive the three-part test used by apex courts to analyze the legitimacy of restrictions to freedom of expression. With these incentives in place for the overremoval and over-moderation of content, we believe the DSA fails this test[8]. **The first report under Article 35 (2) and the best practices therein contained should take this into account and refrain from identifying more risks than those already existent in the law until there are certainties around the concept of "systemic risk" in ways that could conflict with freedom of expression.**

State action remains one of the most prominent risks to Fundamental Rights under the DSA. This was highlighted by Wikimedia in their risk assessment and mitigation report:

> *"Other risks facing the Wikimedia movement may be as important or greater than those reflected in this particular document: for example state surveillance, state censorship, Strategic Litigation Against Public Participation (SLAPPs), and ensuring equity among global minorities. The Foundation is actively working on these, as well as EU DSA systemic risks, and the focus/priority ratings in the SRAM Register should take that rival allocation of resources into account."*.

In a similar fashion, Google stated that they have internal checks against arbitrary state action in government-issued removal requests:

> *"We assess the legitimacy and completeness of government requests, which must be in writing, as specific possible about the content to be removed, and clear in their explanation of how the content is illegal. We do not honour requests that have not been made through appropriate channels. If we receive an oral request, we ask for it in writing. In some narrow cases, to protect the rights of users, we do not act on orders that appear illegitimate or inapplicable"*.[9]

In a forthcoming policy paper, we argue that platforms should document any action by the state that could entail collateral censorship or undue pressure[10], make it public, and include a section in their risk assessment and mitigation reports dealing with how the

[8] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez-Ugarte, Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation (February 28, 2025). CELE Research Paper No. 64 (Forthcoming in the Journal of Intellectual Property, Information Technology and Electronic Commerce, vol. 16, No. 2 (2025), Available at SSRN: https://ssrn.com/abstract=5161173 or http://dx.doi.org/10.2139/ssrn.5161173

[9] Google Ireland Limited, Report of Systemic Risk Assessments 2024, p. 26

[10] For more on this, see CELE's symposium on regulatory threats, especially the pieces by Joan Barata, Paddy Leerssen and Daphne Keller, available at https://observatoriolegislativocele.com/tag/simposio_amenazas/

State could be a risk for Fundamental Rights. **The guidelines in Article 35 (2) report should make sure companies report on the threats to freedom of expression that stem from their contacts with national authorities as a result of the DSA, including but not limited to Article 9 orders.**

Generally speaking, the risk assessment and mitigation reports issued by VLOPs/VLOSEs offer little to no data to substantiate most of their claims. This results in hollow assessments filled with unverifiable assertions. The analysis published by the DSA Civil Society Coordination Group report is explicit on this point: "the contextual information and data that accompanies descriptions of mitigation measures across all platforms analysed are abstract and for the most part unsubstantiated by either qualitative or quantitative evidence, thereby rendering the claims baseless".[11] The analysis issued by the Civil Society Coordination Group also provides numerous examples of the types of information that should have been included to assess the effectiveness of different mitigation measures.[12] Moreover, the reports fail to include any figures or methodological explanations to justify the risk severity assessments made by companies (as required by Article 34.1 of the DSA), or to support the calculation of the so-called "residual risk" after mitigation measures were applied. The analysis published by the Civil Society Coordination Group also outlines the fact that VLOPS have assessed risks mostly as dependent on the conduct of end users, while overlooking the design and affordances of platforms.

All in all, most assessments look like what they actually are: reports produced by companies to present their own operations in the best possible light, avoid fines, and comply with the law while giving away as little information as possible. This is rational business conduct, made possible by the lack of certainty of obligations under the DSA.

The first report under Article 35 (2) should insist upon the lack of data context in support of most of the companies' assertions in their systemic risk assessment and mitigation reports.

[11] DSA Civil Society Coordination Group, "Initial Analysis on the First Round of Risk Asessments Reports under the EU Digital Services Act", march 2025, available at https://cdt.org/wp-content/uploads/2025/03/RA-Report-Assessment-Report.pdf , p. 15

[12] DSA Civil Society Coordination Group, "Initial Analysis on the First Round of Risk Asessments Reports under the EU Digital Services Act", march 2025, available at https://cdt.org/wp-content/uploads/2025/03/RA-Report-Assessment-Report.pdf , p. 16-20

2. QUESTION 2. The report to be published once a year by the European Board of Digital Services in cooperation with the Commission pursuant to Article 35(2) DSA should indicate best practices for mitigation measures implemented by the providers of VLOPs and VLOSEs.

2.A. Please provide examples of practices addressing any systemic risks you have identified, specifying to which systemic risks such measures relate.

2.B. Please refer to any existing documentation, research or resources that could help substantiate the information on the risk mitigation practices you refer to.

Response to question 2:

In their risk assessment and mitigation reports, VLOPs/VLOSEs do not appropriately distinguish between the risk mitigation measures they took against legal vs. illegal content, nor between those taken voluntarily *vis à vis* those taken as a consequence of a legal mandate. These are extremely relevant criteria because, as already stated in the answer to the previous question, the State cannot lawfully require private parties to take action against legal speech based on its content[13], as that would amount to collateral censorship: "when A censors B out of fear that the government will hold A liable for the effects of B's speech"[14]. This does not mean that platforms cannot moderate legal content; however, one thing is for them to do so voluntarily under the immunities granted to them by the law, and another very different thing is for them to act under a legal mandate to "mitigate" certain legal expression. While the DSA's "good samaritan" clause allows the former, the latter constitutes a public interference on freedom of expression and triggers heightened scrutiny under International Human Rights and European Fundamental Rights law, which the DSA fails[15]. **The best practices and recommended measures issued under Article 35 (2) must not include any recommendation of acting upon any expression based on its content under any criteria other than its legality.**

---

[13] See Martin Husovec, The Digital Services Act's red line: what the Commission can and cannot do about disinformation. Journal of Media Law, 16(1), 47–56. https://doi.org/10.1080/17577632.2024.2362483

[14] See Jack Balkin, "Free Speech and Hostile Environments," Columbia Law Review 99, no. 8 (December 1999): 2296.

[15] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez-Ugarte, Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation (February 28, 2025). CELE Research Paper No. 64 (Forthcoming in the Journal of Intellectual Property, Information Technology and Electronic Commerce, vol. 16, No. 2 (2025), Available at SSRN: https://ssrn.com/abstract=5161173 or http://dx.doi.org/10.2139/ssrn.5161173

3. QUESTION 3. When conducting risk assessments, according to Article 34 (2) DSA, providers of VLOPs and VLOSEs must take into account how the identified systemic risks are influenced by risk factors, such as recommender systems and other algorithmic systems, advertising systems, and content moderation systems. The assessment must consider how the risks are influenced by intentional manipulation of the service, including by inauthentic use or exploitation as well as the amplification and potentially rapid and wide dissemination of illegal content and of information that is incompatible with their terms and conditions. The assessment shall take into account specific regional or linguistic aspects, including when specific to a Member State.

3.A. Please provide any information you have of the influence of these risk factors on the systemic risks you have identified.

3.B. Please specify the risk factors and the systemic risks concerned and refer to any existing documentation, research or resources that could help substantiate the evidence you provide.

Response to question 3:

Historically, volume, speed, and accessibility have not been relevant to determine whether a particular piece of expression is legal or illegal but rather to assess the nature and amount of the damages potentially awarded to the victims of the legally reprehensible content. Consequently, only after a court determines that a piece of speech is harmful—and that the harm is both legally cognizable and attributable—it may then evaluate factors such as volume, speed, or permanence to determine the appropriate level of damages. In defamation cases, for instance, advocates and judges would first need to establish the defamatory nature of the statements. Only then would they look at the accessibility and reach of such statements to assess damages. The volume or replication of that speech would then serve as a basis to calculate the compensation owed to the injured party. However, neither speed nor accessibility would, in themselves, render the statement defamatory.[16]

The previous paragraph is still an accurate depiction of the way the law works. Our modern legal systems do not contemplate that certain legal expressions becoming illegal

[16] Agustina Del Campo, Volume, Speed, and Accessibility as Autonomous Harms: Can Modern Legal Systems Deal With Harmful but Legal Content?, at Steven Feldstein (ed.), New Digital Dilemmas: Resisting Autocrats, Navigating Geopolitics, Confronting Platforms, Carnegie Endowment, 2023, available at https://carnegieendowment.org/research/2023/11/new-digital-dilemmas-resisting-autocrats-navigating-geopolitics-confronting-platforms?lang=en

based on the number of times a certain piece is shared or based on whether a certain statement can remain accessible through time. For Robert Post,

> *"the scale of the internet produces forms of harm that may best be characterized as stochastic. Previously, we asked whether particular speech acts might cause harm. The internet has rendered this kind of question almost obsolete. Speech that is simultaneously distributed to billions of persons may produce harm in ways that cannot meaningfully be conceptualized through the lens of discreet causality. We will need instead to think in terms of the statistical probability of harm. Yet at present we lack any legal framework capable of assessing such stochastic harms in ways that will not drastically over-regulate speech".*[17]

The law has been struggling to provide adequate solutions the challenges posed by speed, scope and permanence for some time now. The so-called "right to be forgotten" has been the answer by the CJEU to the problem of the indefinite permanence of information online. It has come at the expense of disastrous consequences for the right of the general public to access information. Bills seeking to enact blanket bans on "disinformation" are, in general, attempts to mitigate the negative effects that certain –otherwise lawful– expressions could have if they viralized really fast and reached a very large number of people. Creating carveouts to internet providers' immunities from liability or forcing them to act upon (or to compulsorily carry) certain legal speech is unlawful under Human Rights law and does not satisfactory solve the difficulties posed by "harmful but legal" content[18].

The DSA should not be interpreted or enforced in such a way that becomes a way to bypass the requirements that International and European Human Rights law places on any state restrictions on freedom of expression. Some of the risk categories specified in the law –those included in Arts. 34.1.c) and 34.1.d)– are broad enough to encompass both legal and illegal expression. Thus, when read together with the mitigation measures in article 35, they entail a duty for platforms to act upon "legal but harmful" expression –e.g., some kinds of legal disinformation–. Besides being at odds with the legality principle under International Human Rights Law, this will likely involve crossing

[17] Post, R., "Democracy and the Internet", Balkinization 20th Anniversary symposium, January 28, 2023. Available at https://balkin.blogspot.com/2023/01/democracy-and-internet.html

[18] Daphne Keller, Lawful but Awful? Control over Legal Speech by Platforms, Governments, and Internet Users (essay), The University of Chicago Law Review, available at https://lawreview.uchicago.edu/online-archive/lawful-awful-control-over-legal-speech-platforms-governments-and-internet-users

Husovec's red line: taking action upon legal speech based on its content[19]. This framework will likely lead to a shrinkage of the European online civic space for the general public. At CELE, we performed an in-depth analysis of how the risk-related obligations of the Digital Services Act conflict with International Human Rights and European Fundamental Rights obligations of member states in our research paper: "Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation"[20], to which we refer for greater clarity on this matter.

We acknowledge the challenges created by the volume of information in the internet, the speed at which it moves and its permanence online. There's some dissonance between these features and the old liability attribution framework, which was created for a different world and could have some aspects revised to account for these phenomena. However, we are convinced that giving up or weakening the right to freedom of expression is not the way out of this stalemate. The illegality of a piece of content should still be the only basis for both attributing liability to speakers and for the State mandating third parties to take action upon it.

At CELE, we believe Human Rights are safeguards rather than dangers to our democracies, and we are committed to working with governments across the globe towards regulations that deal with the harms provoked by human expression, as long as they respect our more basic common normative commitments. **The report issued under Article 35 (2) can take speed, volume, and permanence as factors that have an influence in the harms caused by human expression. However, it should not include any recommendation for platforms to intervene in the public debate in any way that contravenes or circumvents well-established international or European standards on freedom of expression.**

[19] See Martin Husovec, The Digital Services Act's red line: what the Commission can and cannot do about disinformation. Journal of Media Law, 16(1), 47–56. https://doi.org/10.1080/17577632.2024.2362483

[20] Agustina Del Campo, Nicolás Zara and Ramiro Álvarez-Ugarte, Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation (February 28, 2025). CELE Research Paper No. 64 (Forthcoming in the Journal of Intellectual Property, Information Technology and Electronic Commerce, vol. 16, No. 2 (2025), Available at SSRN: https://ssrn.com/abstract=5161173 or http://dx.doi.org/10.2139/ssrn.5161173