

Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks

Agustina Del Campo, Nicolás Zara, y Ramiro Álvarez Ugarte

May 20, 2025

Agustina Del Campo, Nicolás Zara, Ramiro Álvarez Ugarte. 2025. "Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks". In *Proceedings of the Fourth European Workshop on Algorithmic Fairness (EWAF'25)*. Proceedings of Machine Learning Research, 12 pages.

Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks

AGUSTINA DEL CAMPO*, CELE, Argentina

NICOLÁS ZARA*, CELE, Argentina

RAMIRO ÁLVAREZ UGARTE*, CELE, Argentina

The Digital Services Act can potentially become a tool for change toward a more rights-abiding, competitive European digital platform ecosystem. However, as it currently stands, it is prone to be misused as a powerful tool for censorship as well as a tool to displace human rights as the backbone of the rule of law in modern democracies. Its risk-based and “new governance” approach to regulation puts rights on the back burner, and, while paying lip service to them, takes them off the center stage. This is problematic, for the incentives of the regulation and the oversight and enforcement mechanisms it establishes may channel the violation of fundamental rights. This paper proposes ways to correct this in the incipient enforcement and interpretative stage of the statute. In particular, we propose to adopt a working definition for “systemic risks”, clarify and narrow the individual risks identified, interpret them in light of human rights standards; take the proportionality principle seriously and refine oversight; allow room for transparency and accountability for the oversight mechanisms, and bring back the State as a potential source of risks to human rights and other paramount values the DSA seeks to protect.

Keywords: Digital Services Act, DSA, regulatory managerialism, systemic risks, Human Rights, Freedom of Expression: Digital Services Act, DSA, regulatory managerialism, systemic risks, Human Rights, Freedom of Expression

Reference Format: Agustina Del Campo, Nicolás Zara, Ramiro Álvarez Ugarte. 2025. Reclaiming Human Rights for Platform Governance: Proposals for Restoring Their Centrality in the Era of Risks. In *Proceedings of the Fourth European Workshop on Algorithmic Fairness (EWAf'25)*. Proceedings of Machine Learning Research, 12 pages.

1. Introduction

The Digital Services Act (DSA) of the European Union offers a new path forward for regulating big technological companies in the business of facilitating, managing, and profiting from user-generated content (UGC) online[52]. It has incipiently emerged as a regulatory model to non-European countries seeking new ways to govern online speech. It specifically intends to strike a balance between freedom and openness and democratic control over technology. Given the information asymmetries between states and companies and in an effort to create legislation capable of evolving with the speed of technology, the DSA resorts to open-ended provisions and broad language to create

**All authors contributed equally to this research.

Authors' Contact Information: Agustina Del Campo, CELE, Buenos Aires, Argentina, adelcampoiso@gmail.com; Nicolás Zara, CELE, Buenos Aires, Argentina, nicolaszara.cele@gmail.com; Ramiro Álvarez Ugarte, CELE, Buenos Aires, Argentina, ramiroau@gmail.com

This paper is published under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC-BY-NC-ND 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

EWAf'25, June 30–July 02, 2025, Eindhoven, NL

© 2025 Copyright held by the owner/author(s).

reporting obligations as well as company-led assessments regarding pre-identified social “risks”. However, this risk-based approach and its managerial regulatory style put the regulation on a collision course with the standard view of human rights developed and nourished throughout the twentieth century. In a recent paper, we discuss this diagnosis at length and conclude that although the risk-based approach may be problematic, many of the difficulties it brings about may be corrected through interpretation in the implementation stage of the law by the European Commission and the different Member States[11]. Identifying these challenges and potential means to address them can be particularly important to protect human rights in Europe, but also to inform the modelling process that the DSA is leading, alerting other countries on what to fix before adopting similar legislation elsewhere. This paper proposes concrete steps that could be taken to bridge that gap, including defining key terms, interpreting and enforcing the DSA’s open-ended provisions through the prism of international human rights law, linking its due diligence obligations to the Business and Human Rights framework, and acknowledging the role of the state as a potential source of risks to freedom of expression. This would strengthen the European Digital Strategy in two ways. Firstly, adequacy with Human Rights standards would still offer much-needed solutions to the stalemate the DSA enforcement is currently at. Secondly, harmonizing the European legislation with human rights commitments would better serve the purposes of the EU to promote the DSA as a model regulation for other countries.

2. The problem

The DSA can potentially lead to a more rights-abiding, competitive European digital platform ecosystem. However, as it currently stands, it is prone to be misused as a powerful tool for censorship[9, 19, 32] as well as a tool to displace human rights as the backbone of the rule of law in modern democracies. The DSA presents a novel approach to content moderation that is centered on risks and is relatively autonomous from fundamental rights standards. It shifts the platform regulation conversation from a rights-centered perspective to one that has risks at its core, wherein rights infringements are presented as one “systemic risk” category among many others. This change in narrative comes with significant consequences. It takes the law to uncharted territories, introducing new logics that lead to unpredictable outcomes.

Despite common claims that the DSA does not contain substantive rules disallowing expression[30, 53], its risk-based approach to content moderation[52:34 and 35] unprecedentedly expands the scope of legally governed expression well beyond the limits of protected speech under human rights law[10]. While it formally leaves all individual moderation decisions within the field of private action, it also obligates platforms to deal with—at least—some legal content. Thus, it would be hard to think of a platform complying with articles 34 and 35 of the DSA without acting upon some legally permissible content. When private action upon third-party posted content is motivated by or directly linked to a state mandate to intervene, such a mandate should be subjected to the scrutiny called for by the three-prong test widely used in Europe and other regions to assess restrictions to freedom of expression[47:35]. The risk-based approach to content moderation of the DSA does not fare well under this standard. The lack of a definition of “systemic risks” and the open texture of the language in Article 34.1 fail the first step of the test, the requirement of legality[13:44]. These broad definitions also make it difficult to assess the extent to which there is a “pressing social need” that justifies the new governance mandates (the legitimate aim requirement). And the absence of safeguards for over-removal in the face of the vagueness of the statute makes it unlikely that these

provisions could be judged to be “narrowly tailored” to assess the pressing needs they are supposed to address. On the contrary, the DSA seems to rely on vagueness as a feature, and not a bug, of its design. This grim picture is completed by the hefty fines legally established for noncompliance, which work—albeit perhaps unintentionally—as incentives for platforms to over-remove content rather than for them to try to strike a “right” balance in content moderation.

The DSA's general approach is novel at least in the field of speech governance. It was not expected to deliver safety, transparency, and freedom from the outset: it was rather conceived as an iterative process that would allow for stakeholders to make some mistakes while going through a reasonably expected learning curve. However, for the European platform regulation framework to work and be sustained over time, its risk-based approach needs to be made compatible with fundamental rights.

3. How to bring rights back to the table

This paper suggests ways to strengthen the DSA and bridge gaps between the legislation and freedom of expression standards contained in multiple international and regional human rights treaties. Although our focus is mostly on freedom of expression, strengthening the interpretation of the DSA through human rights law can also help strengthen other human rights along the way, including the rights of minors and adolescents, the rights of women, due process, and the rule of law.

Human rights law has been developed over the last 80 years and has been a guiding force in international relations, legislation, and jurisprudence in different jurisdictions. Despite its ambiguities, it has provided a common set of principles and a common language among nations. Human Rights norms set the minimum common standards that every state must afford to every person within its jurisdiction and control. The language of the treaties has been developed and has evolved through time with their adoption, their local implementation, and the ensuing dialogue between legislatures, tribunals, legislators, international organizations, and supervisory bodies. Leaning and building on this legal corpus provides many benefits, including the adoption of existing standards over the invention of new ones, taking advantage of already tested expertise and legal knowledge, supporting the legislation through a common language and showing in the process adherence and respect to human dignity as defined by the international community over time.

The suggestions we make in this paper are not particularly creative but rather obvious under the human rights standards that we cite and believe should be the starting point in any platform regulation effort. They include clearly defining what should be understood by systemic risks; clearly defining what risks are companies expected to address and tying those risks to existing human rights law; committing to clear and express proportionality benchmarks; include state action as a factor impacting risks to human rights within companies; and committing to transparency and accountability in the implementation of the DSA.

3.1 Defining systemic risks

When confronted with the task of defining “obscenity” as a category of unprotected speech under the First Amendment, Justice Potter Stewart famously wrote: “I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description, and perhaps I could never succeed in intelligibly doing so. But I know it when I see it”[45]. DSA stakeholders seem very much in line with Justice Stewart when working with systemic risks under the DSA.X

The European legislator created a statute that is imprecise by design. The degree of flexibility it provides is intended to foster a “regulatory dialogue” between VLOPs/VLOSEs and the European Commission, where the former have the flexibility to assess and mitigate the risks present in their systems as they see fit [10] as long as they periodically report such efforts in legally mandated risk and transparency reports. The Commission learns from the process, gathers all sorts of evidence and information, and incrementally provides feedback to them in the form of best practices and delegated regulation. This process is also informed by the input of civil society, academia, and the people at large through consultations, periodic feedback, and diverse forms of engagement. This design feature explains the absence of a definition of what a “systemic risk” is in the DSA and the European Commission’s decision not to provide one.

Additionally, some of the (non-exhaustive) risk categories of Article 34 are vague and nonspecific in their wording. This open-ended text grants both platforms and enforcement authorities great discretion, increasing the chances of both regulatory capture by platforms and undue pressure on them from enforcement bodies.

Scholars[36] and civil society organizations[3, 26] have long advocated for clearer definitions. It is now widely accepted that “much clarity on the meaning of systemic risks is needed in order to provide legal certainty, avoid evasion of the legal rules by the responsible platforms and search engines, and ultimately ensure sound protections of EU citizens online”[20]. However, the concept of systemic risks remains obscure[28] given its own vagueness and the lack of guidance by authorities. This is particularly relevant because VLOPs/VLOSEs are expected to identify any and all systemic risks that they are exposed to, assess and mitigate them. These concerns are shared across the spectrum of DSA stakeholders[16].

Despite claiming to regulate only processes, the risk management clauses in Article 34 and the mitigation measures in Article 35, when read together, impose content moderation obligations. In some cases, these obligations entail the duty to *act upon* (remove, de-amplify, demonetize, filter, de-rank, cease to suggest, sort, geoblock, etc.) certain expressions that are legally protected: the so-called *legal but harmful* speech [29]. Of course, in the pre-DSA world, platforms used to moderate this kind of content anyway. However, one thing is to do so voluntarily, and another very different is to act under a legal mandate to “mitigate” potential harms created by legal expression. The latter constitutes a public interference on freedom of expression and therefore triggers a heightened level of scrutiny under international and European human rights standards[29:256], which—as will be developed in the next subsection—the DSA does not pass.

The principle of legality serves as a safeguard against excessive discretion and arbitrariness in the enforcement of imprecise regulation.[46] Ideally, the text of the DSA should be revised to clarify and define what constitutes systemic risks that companies should mitigate. This would limit the scope of the DSA for both companies and states. Although it would potentially limit the ability of the European Commission (EC) to address new or future risks related to technology, it would allow better enforcement and actual accountability both of the platforms and the Commission’s members and European states in the implementation. It would also ease anxieties regarding potential shifts and turns in the future politics of the EC and the EU.

Still, the DSA is rather new, and it has not been fully implemented yet. It is therefore not likely that the EU Parliament will introduce changes to its text shortly. Instead, the European Commission is responsible for establishing and disseminating guidelines for the Act’s enforcement, a process that has been underway since 2022. The Commission has issued Delegated Acts that inform the implementation of the DSA and contribute to clarifying certain aspects of it that were left too abstract or too broad within the Act itself. In pursuing this path, it is imperative that the

European Commission shares a working definition of systemic risks so that every other stakeholder within the ecosystem may fully understand what they are looking for and after.

The difference between Justice Stewart and companies operating under the DSA is that the former was a State actor passing judgment on an individual case. Without clear definitions and guidelines, the State cannot mandate companies to assess individual cases or to take care of certain risks, as outlined in the DSA, while exposing them to the risk of being heavily fined in case they fail to properly mitigate them.

Clearly defining what a systemic risk is under the DSA is imperative for making the act compliant with human rights standards. This clarity is also necessary for the Act's coherence and to alleviate concerns about potential abuse in the future.

3.2 Fixing vagueness, one systemic risk at a time

The DSA, as it currently stands, nudges platforms towards classifying and curating content in ways that violate the principle of legality embedded in the three-prong test traditionally used to assess restrictions to freedom of expression. This is not only because the category of systemic risk is not defined, but also because the individual instances of risks to be addressed are also described in vague and abstract terms that fail to provide “guidance for action”. Article 34.1 of the DSA classifies risks into: (a) dissemination of illegal content, (b) negative effects for the exercise of fundamental rights, (c) negative effects on civic discourse, electoral processes, and public security, and (d) negative effects in relation to gender-based violence, the protection of public health and minors and serious negative consequences to the person’s physical and mental well-being. It also provides that those risks can stem either from the platform’s system or the use made of them (this includes third-party posted content). Some of the mitigation obligations in Article 35 are to be applied to content deemed risky under Article 34.1 categories, even when most of the speech that falls under 34.1(c) and (d) is perfectly lawful. That places the DSA on a collision course with International and European Human Rights and Fundamental Rights law.

When confronted with the legality principle of the three-part test, the risk assessment and mitigation provisions in the DSA fail. The European Convention on Human Rights mandates that any interference with freedom of expression must be prescribed by law [54:10.2]. Under the European Court of Human Rights case law, this demands that such law be accessible—published and publicly available—and formulated with sufficient precision to enable the citizens to regulate their conduct: they “must be able—if need be with appropriate advice—to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail”[46, 50]. The ICCPR adopts a similar standard [49]. While the European case law does not require absolute certainty and allows for some flexibility[48], laws restricting freedom of expression “may not confer unfettered discretion on those charged with their execution”. [46] However, as it currently stands, the DSA does exactly that: the imprecise categories of “risky” content it creates grant private platforms enormous discretion in defining and enforcing content restrictions.

Professor Husovec warned that careful enforcement of what he called the DSA “red lines” is necessary to prevent it from infringing upon freedom of expression rights. He argued that VLOPs and VLOSEs must ensure they employ only content-neutral measures to mitigate risks posed by legal content, while content-based restrictions could be deployed against illegal expression[23:53]. His point, however, could be partially disputed. Generally, the state can single out certain legal activities as undesirable and take steps to discourage them if their negative consequences outweigh their benefits. For instance, a state can replace car lanes with cycling lanes or raise taxes for car ownership

due to environmental, health, and road safety considerations. It can ban single-use plastics. Nobody has a fundamental right to commute by car quickly or to use plastic straws. The analysis is different, however, when the disfavored activity is the free enjoyment of a human right, such as lawfully expressing oneself or accessing information. Such a measure ought to undergo stricter, more skeptical legal scrutiny. The state has the burden of showing very good reasons to require a third party to design a product or service in such a way that discourages the regular exercise of a fundamental right (i.e., user's rights to express and access information), for it could be illegitimately impinging on its exercise using that third party as a proxy. Therefore, legal content, even if annoying, preposterous or unpopular, shocking or disturbing, should be let to flow freely[13:49].

Having said that, we can agree on one baseline point: Husovec is right in saying that the EC cannot make platforms apply content-based mitigation measures against legal though “risky” expression. Yet, we have already seen action by the European Commission pointing to the duties of platforms to act upon legal content. Everybody remembers Commissioner Thierry Bretton going after platforms for spreading “disinformation” in connection with the Israel-Gaza conflict[39] or admonishing Elon Musk before he broadcasted a call with then-presidential candidate Trump by stating that the Commission was monitoring “potential risks in the EU associated with the dissemination of content that may incite violence, hate and racism in conjunction with major political---or societal---events around the world, including debates and interviews in the context of elections”[40]. Unfortunately, this is not only a tale of a rogue commissioner. The Guidelines for providers of VLOPs and VLOSEs on the mitigation of systemic risks for electoral processes[51] require platforms to have a tighter grip on expression during electoral periods and to take action in connection with certain pieces of content, regardless of their lawfulness, such as disinformation, extremist, and radicalizing content[51:27.g]. The Commission nudged platforms into doing so by stating that VLOPs and VLOSEs that “do not follow these guidelines must prove to the Commission that the measures undertaken are equally effective in mitigating the risks” [55].

More recently, the Codes of Practice on Disinformation[56] and Hate Speech[57] have been incorporated as Codes of Conduct under Article 45 of the DSA. The voluntary nature of the Codes of Practice, which precede the DSA, has already been questioned in various instances[31, 43]. But even assuming their voluntariness, the mere participation of the European Commission in the drafting process and the fact that, once incorporated, they constitute a benchmark for DSA implementation led some authors to conclude that they should not include “content-specific (Key performance indicators) KPIs linked to lawful content at all”[23:54]. In spite of that opinion, the Code on Disinformation includes commitments to demonetize disinformation without distinguishing between legal and illegal disinformation. For instance, under Commitment no. 1 (“Demonetisation of disinformation”), signatories pledge to “defund the dissemination of disinformation”, to deploy policies to avoid the “publishing and carriage of harmful disinformation”, and to “take steps to avoid the placement of advertising next to disinformation content”[14]. In addition, the Code of Conduct on Illegal Hate Speech lacks a concrete definition for hate speech and includes metrics that encourage over-removal of content, such as the 24-hour deadline for takedowns and the commitment to review at least 50% of the notices received[25].

While the lack of substantive guidance on the meaning of “systemic risk” generates uncertainties in identifying unenumerated risks, the absence of interpretive parameters for risk categories makes the specific risk categories in article 34 (1) ethereal, almost unassailable. This was shown by the first round of risk assessment and mitigation reports published in November 2024. Using a plausible interpretation of the broad latitude given by the law to them in identifying, assessing, and mitigating risks, companies have chosen to fulfil their reporting duties using a very high

level of abstraction and without saying much. For the most part, companies have reported their existing trust and safety policies as mitigation measures adopted per DSA compliance.

The vagueness in the legal definitions of risk categories, the European Commission’s past conduct and the fines for noncompliance prescribed by law strongly incentivize platforms towards interpreting risks broadly, and mitigating them by adjusting their terms of service and enforcement policies to act upon all kinds of legal and “borderline” content, thus dramatically reducing online civic space. The Office of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression has warned against states that “impose obligations on companies to restrict content under vague or complex legal criteria without prior judicial review and with the threat of harsh penalties”[18].

Leerssen argues that the risk management obligations under the DSA are so vague and flexible that the scope of the European Commission’s authority in its enforcement is almost impossible to discern[27]. That needs to change. Some categories in the DSA are, in that sense, less problematic, such as those that require assessing risks of dissemination of illegal content and negative effects on fundamental rights[52:34.1.a and 34.1.b]. These are arguably legitimate¹ categories insofar as they do not touch beyond what the state could lawfully prohibit or mandate acting by itself and are attached to preexistent legal frameworks—domestic and European law in the case of 34.1.a. and fundamental rights law in the case of 34.1.b. That is not the case, however, with the risk categories in Article 34.1.c and 34.1.cd. Not only it is unclear what amounts to a negative effect on civic discourse, electoral processes or public security, nor is there any certainty as to what exactly is a negative effect in connection with gender-based violence, the protection of public health and minors, or the person’s physical and mental well-being[52:34.1.c and 34.1.d], but also many of the expressions that could be reasonably classified under these categories are perfectly lawful. If that is the case, then it is illegal for the State to use platforms as a proxy to act on content it could not regulate directly[1:28].

3.3 Reading Human Rights Standards into the text of Article 34 of the DSA

Given the ambiguity of the existing legislation and the lack of conformity to human rights standards, the most reasonable and legally sound path forward should be to interpret and enforce risk categories in the light of well-established legal concepts, especially where relevant Human Rights standards are available². For instance, when Article 34.1.d refers to “the protection of minors”, this should be understood as a reference to the text of the International Convention on the Rights of the Child, the interpretations and recommendations of the Committee on the Rights of the Child (CRC), and other provisions of any relevant Human Rights and Fundamental Rights treaties (such as Article 24 of the Charter, as interpreted by the CJEU) to which member states are parties.

The benefits of this approach are twofold. On the one hand, it would contribute to mitigating the problem previously described in connection with the principle of legality, bridging the gap between the DSA’s risk-based approach and Fundamental Rights and Human Rights Law. On the other hand, grounding the content of these concrete obligations in the duty of the European Union and its member states—under both European and international law—to protect human rights, would enhance their legitimacy, particularly at a time when the broader structure of the European regulatory project is under external attack.

¹ In the case of obligations to act upon illegal content, whether or not they are legitimate in fact is a different question, which answer rests on the compatibility of domestic laws with constitutional and conventional standards.

² We side-step here the issue of whether the vagueness of the provisions that define “systemic risks” in the DSA can be solved solely through acts of interpretation and enforcement. While we are aware that an actual revision is unlikely, the legally-sound and generally accepted solution to a vague legislative provision is a precise legislative provision.

Anchoring the DSA's risk assessment and mitigation obligations—now resisted by the same platforms that proudly boasted about their lengthy reports only a few months ago—in Human Rights Law would enable the shaky theoretical foundations of the DSA to stand on firm ground. Backed by 80 years of International Human Rights Law development through treaties, case law, customary international law and scholarly work, it could remain open and flexible to incorporate future developments in the technology field.

This does not entail abandoning the risk management obligations or the meta-regulation scheme altogether in favor of a classic command-and-control kind of legislation. As we imagine it, this shift would make the obligations under Articles 34 and 35 of the DSA much more akin to those under the United Nations Guiding Principles on Business and Human Rights (UNGPs). However, unlike the UNGPs, DSA obligations would be industry-specific, mandatory, and supported by oversight and enforcement mechanisms. This is hardly an implausible idea considering that the DSA was, to a great extent, inspired by and borrowed language from the UNGPs[10].

The endeavor will probably face both state and civil society resistance, though. The risk-based approach of the DSA, in its current form, is embraced by advocates of many diverse causes. Its broad language provides radically different movements with a fertile ground to express their policy preferences as pressing public values[35:11]. These activists might perceive the strict rules of International Human Rights Law as a nuisance that separates them from their laudable policy goals, and could feel uncomfortable with the prospect of Freedom of Expression often prevailing against some of these competing interests (information integrity, safety, the well-being of children, the health of the public conversation, algorithmic fairness, national security, electoral integrity, and so on). The European Commission, in turn, benefits from these interpretive disputes, for they broaden the scope of its enforcement authority with each new expansion in the interpretation of these risks. With few proportionality safeguards in place—as we will see in the next subsection—the scope of the risks to be managed under the DSA is likely to grow in many directions.

As obvious as it may seem for some, it is critical to emphasize that strictly aligning the content of the DSA's obligations with Human Rights and Fundamental Rights would do a lot to protect today's citizens from any overreach of the current European Commission and, most importantly, tomorrow's citizens from potential bad-faith action by a potentially authoritarian Commission in the future. At the same time, it would reduce the chances of regulatory capture by companies[38] through the deployment of discretionary, self-serving conceptions of "risks". When rights are treated as only one risk among many competing considerations, balancing necessarily becomes arbitrary and subjective. Bringing rights back to the center of the conversation makes balancing possible and provides for easier and more predictable enforcement.

The incorporation of International Human Rights into the interpretation of the DSA could also be the key to its survival and to its potential as a model legislation. On the one hand, the European Digital Strategy in general and the DSA in particular are currently under fire from the most powerful American VLOPs/VLOSEs, with full backing from the US government[8, 34, 41, 42]. Anchoring interpretation and enforcement of the DSA in International Human Rights standards would shield it against growing accusations of censorship. On the other hand, the potential for the DSA to become the golden standard in platform regulations and a model after which many bills are drafted is also threatened if not closely aligned with Human Rights standards.

While the incorporation of this body of doctrine into the DSA would be a significant improvement and a promising first step, it would by no means automatically solve all problems. As a matter of fact, human rights are, by definition, vague and sometimes paradoxical[33:7], and therefore, the translation of general principles into particular rules remains a challenge[37:969]. That being said, interpreting the DSA through the lens of Fundamental Rights and

Human Rights Law would mean incorporating a set of shared consensus and a tradition, which would provide a much needed “common conceptual language”[37:967] that could be of use to all stakeholders, without which it is now virtually impossible to explain, contest, or argue for or against State or platform’s decisions. The common language of Human Rights allows for debates and deliberation[12:47] in a way that the risk framework does not, since it is not attached to any meaningful set of shared substantive commitments.

3.4 Fixing proportionality

Proportionality is not completely absent from the DSA. The asymmetric obligations for different kinds and sizes of platforms are a step in the right direction³. Moreover, internal complaints and appeals systems, as well as out-of-court dispute settlement systems, give individual users the chance to challenge potential over-removal of content and therefore protect their infringed upon rights[52:20 and 21]. However, the systemic risk-related portions of the DSA do not withstand scrutiny under this last step of the three-part test. When the scope of obligations is not clear, as shown in the previous subsections, there can be no proportionate measures or—to put it less bluntly—proportionality becomes impossible to assess.

The problem is linked to the lack of a precise benchmark to use in the context of a proportionality analysis. Although Article 35 demands proportionality in risk mitigation, the law contains no concrete safeguards against “disproportionate mitigation” of risks nor any guidance as to the level of “residual”⁴ risk that could be considered welcome, tolerable, or unacceptable. Should proportionality be measured against the size of the risk? Or against the level of restriction of the right?

The lack of a benchmark could, in theory, be solved through practice. But in its current form, the DSA prescribes that auditors are the only stakeholders with enough access to data and technical capacity to evaluate the proportionality of mitigation measures. It is unlikely, however, that they will ever find mitigation measures disproportionate, for a number of reasons. First, the DSA does not require auditors to have any expertise on Human Rights but solely on risk management[52:37.3]. By design, auditors have no expertise on human rights and their practice will not, as a consequence, be capable of filling the gaps left by an ill-designed portion of the law. Second, and even though the instructions set out in articles 13 and 14 of the Delegated Act on Independent Audits mandate them to assess proportionality, reasonableness, and effectiveness, the indicators identified therein focus only on the last two. Third, and finally, auditors are too private and too dependent on corporations to satisfy the value of independence the DSA declares in any meaningful sense. This is a further reason to doubt they will be capable of filling the benchmark gaps left open by the law regarding proportionality.

In addition to the work of auditors, the European Commission oversees the adequacy of VLOPs/VLOSEs’ activities with their obligations under the DSA. It can initiate investigations, request information, issue best practices jointly with the Board of Digital Services, and ultimately impose onerous fines that can add up to 6% of their total annual turnover[52:74.1]. So far, we have not seen any enforcement action related to the over-removal of content or overenforcement of the DSA. The first year of enforcement efforts has rather gone in the opposite direction. For instance, the European Commission has initiated proceedings against X for failing to remove “misinformation”, a

³ But *see* [17]

⁴ This is the term used by platforms in their risk reports to refer to the level of risk that remains once mitigation measures have been applied.

category not included in the DSA and that includes both legal and illegal content[58]. These procedures, of which we learned about solely through press releases, could be turned into a jawboning tool for the EC against platforms[10].

As a result, when faced to the task of implementing the imprecise language, platforms' incentives become clear: it is safer to incur in over-moderation invoking the "good samaritan clause" of the DSA [52:7] than to leave any risk "insufficiently mitigated", and be found noncompliant and pay a huge fine. This can hugely affect the users' rights to access and share lawful expression online, dramatically reducing the online civic space. At the same time, it makes the DSA an ideal tool for collateral censorship---"when A censors B out of fear that the government will hold A liable for the effects of B's speech"[4:2296].

In sum, the independent auditors foreseen by the DSA are ill-suited to control the proportionality of the mitigation measures taken by platforms because they are not human rights experts and lack the institutional infrastructure and legitimacy of courts. The European Commission seems not to be concerned with the problem of over-removal of lawful content at all. And—to make matters worst—researchers and civil society organizations do not have enough access to platform data to weigh in and check on auditors' actions, notwithstanding the DSA's provisions on the contrary. Until now, at least, the DSA has failed to deliver on this promise. In this scenario, nobody is there to police Husovec's red line, and the human rights due diligence obligations under articles 34 and 35 are at risk of becoming checkbox compliance exercises by companies, later certified by auditors, with the European Commission looking in the direction of other priorities.

There's no easy fix for this problem, but we suggest a few steps that could be taken in that direction. First, the legality principle in the DSA must be fixed: there cannot be any proportional mitigation measures if the scope of the risks is nebulous. Furthermore, the Delegated Act on Independent Audits should incorporate granularity and KPIs capable of helping different stakeholders in assessing proportionality. And proportionality should be clearly understood as related to the conflicting rights restrictions rather than the potential magnitude of the undefined social risk. Risk assessment and mitigation reports should include explanations by VLOPs/VLOSEs showing why the mitigation measures chosen are less speech-restrictive than alternative ones. Platforms should also demonstrate that they deal with legal and illegal content differently.

External audits are not well-equipped to evaluate proportionality, and this is a fundamental weakness of the framework. As long as they remain the main oversight mechanism, unless risk mitigation reports include the underlying data in support of their conclusions, and until scholars and Civil Society Organizations get enough access to platform data to assume the watchdog function that the Act assigns them, controlling proportionality in risk mitigation under the DSA will remain a black box and platform accountability will remain chimeric.

3.5 Facing the public

The DSA envisions a continuous regulatory dialogue in which platforms and regulators learn from each other. The theory behind the act is that regulators would draw conclusions from platform reports, information obtained through their investigative powers, and insights from other DSA stakeholders, incorporating them into industry best practices and standards that would, in turn, help refine systemic risk categories. VLOPs/VLOSEs would then implement these incremental measures and best practices guidelines prescribed by regulators.

However, as a precondition to the existence of such a dialogue and to prevent the possibility of state overreach, all stages of this process must be as open and transparent as possible. The DSA establishes many engagement mechanisms between platforms and state actors[5], which could turn into regulatory pressure (*jawboning*) unless

transparency is guaranteed throughout all stages of implementation. In that sense, platforms must make every effort available to substantiate the key points in their risk assessment and mitigation reports with relevant data, thereby enabling public debate and scrutiny. In the first round of reports, for instance, VLOPs/VLOSEs failed to justify the criteria used to determine the severity of risks and the level of residual risk remaining after mitigation measures were applied.

Similarly, the European Commission must ensure transparency in proceedings against platforms by publishing as much information as possible, beyond mere press releases as it currently does. This is essential not only for access to information but also to legitimize these proceedings and overcome accusations of undue pressure on platforms. The backlash against Thierry Breton's open pit pressuring campaign against platforms highlights the importance of full transparency and (at the same time, and paradoxically) suggest that regulators incentives may lay somewhere else[2].

Meaningful engagement by both platforms and state actors is mandated by the DSA[52 recital 90]. Without a properly working environment where all stakeholders inform one another's work, the system is bound to fail[21]. Civil society and academia hold valuable expertise that could enhance risk assessment and mitigation reports, yet their inputs remain largely overlooked and this is a key weakness in the dialogue the DSA seeks to foster. Without the involvement of the public, the dialogue between platforms and regulators becomes an exercise in technocratic governance, bound to be perceived—especially these days—as illegitimate and driven by powerful elites. To make dialogue meaningful, the public must also be brought back in. There are many innovations that European and national regulators could adopt to make this inclusion meaningful and the DSA does not prevent them. There is vast room for improvement.

3.6 Checking on state actors too

The DSA includes no specific safeguards against state overreach. Some would say that placing the burden of controlling VLOPs/VLOSEs on the European Commission is a safeguard against the less-trustworthy member states in the Union. However, even in that case, the Commission's power comes virtually unchecked. While the DSA considers platforms and their users as potential risk generators, state actors are not mentioned. Platforms should document any state action that they perceive as a potential threat to human rights, make it public, and inform it in their risk assessment reports, along with the mitigation measures adopted whenever possible. Failing to acknowledge the role of the state as a potential risk-generator creates a fundamentally incomplete picture of the risks to our human rights online. To comply with international Human Rights law, any definition of systemic risk needs to acknowledge that risks can emanate from both private and public actors. Interestingly, both Wikimedia and Google have included states as sources of risk in their first reports in 2024. As demonstrated in this piece, the state remains a significant risk for freedom of expression through collateral censorship mechanisms.

4 Conclusion

The DSA is a unique and experimental piece of legislation that attempts to bring much needed accountability to tech companies, particularly those managing user generated content. Its drafting was followed by companies, States, civil society and academia with great interest and the resulting legislation is currently starting to be implemented and

interpreted. This paper outlines a series of challenges that the drafting of this legislation brought about and that, we argue, can be fixed through its interpretation and implementation.

In their ambition to create a catch-all piece of legislation that allows to reign in the harms produced by large internet companies, the EU Parliament and the Council have created an extremely vague and sometimes overbroad piece of legislation that unintendedly undermines the existing rule of law, threatens the full exercise of human rights online and has “significant potential for state censorship”[32]. In the quest for flexibility to adapt to an ever-changing environment, they devised a legal instrument that awards enormous swaths of discretion to both regulators and companies to interpret the meaning and scope of key concepts that the law leaves undefined.

We believe that aligning this regime as closely with International Human Rights standards as possible would fix some of the most salient challenges raised against the legislation, ease its implementation, and provide the much-needed legal stability for every stakeholder involved.

The DSA is more than the European Union’s flagship platform regulation statute. It is the endpoint of a multi-year struggle to grapple with some of the most pressing challenges of our time. It is a law that marks the end of an era⁵ and captures what could be the beginning of a new paradigm. Dubbed by some as a “digital civil charter” [22:912] or even a “constitution for the Internet”[44], it embodies the European idea of a rights-driven regulatory model for the Internet [7]. If Europe is serious about making the DSA a standard-bearer legislation for the whole world, its framework should be further shielded against state overreach and private regulatory capture, or else its risk-based approach will sooner or later be used to silence dissident voices by both governments and platforms alike [24], either inside the European Union or outside.

The diverse nature of these challenges makes a single solution for all of them problematic. Much like the printing press, the Internet has initiated a revolution in terms of the speed of circulation, the scale, and the permanence of expression. But not all internet evils are harmful in a legally relevant sense, and not all those relevant harms are legally redressable. Human Rights Law demands that we tolerate some of the harms that the spread of some ideas we don’t like can inflict on our societies. The value of those standards and their soundness in today’s digital environment could be discussed, but that is outside of the scope of this piece. That being said, unless those standards change, they must be respected. We believe there are, and we have argued for, some good reasons to do so.

Recalibrating our expectations is crucial: the DSA is no magician’s hat. Maybe it cannot single-handedly save Europe or the world from the rise of authoritarian governments[6, 15], but if carefully interpreted and transparently implemented, it can provide some of the long-awaited platform accountability that Europe has been striving for and come in handy for the protection of our most fundamental rights online. This paper proposes 6 concrete steps in that direction: adopt a working definition for “systemic risks”, clarify and narrow the individual risks identified, interpret them in light of human rights standards; take the proportionality principle seriously and refine oversight; allow room for transparency and accountability for the oversight mechanisms, and bring back the State as a potential source of risks to human rights and other paramount values the DSA seeks to protect.

⁵ Epitomised also by the many amendment and carveout proposals for Section 230 of the Communications Decency Act, coming from both sides of the aisle in the United State Congress.

Acknowledgements

We would like to thank our colleagues at CELE for helpful comments to a previous version. Feedback from reviewers was highly welcomed. The errors that remain are our responsibility.

4. References

- [1] 9th Cir. 2025. *NetChoice v. Bonta*.
- [2] Access Now, Article 19, and Electronic Frontiers Foundation. 2024. Civil society statement: Commissioner Breton needs to stop politicising the Digital Services Act. Retrieved from <https://www.article19.org/wp-content/uploads/2024/08/Letter-to-Commissioner-Breton-August-2024.pdf>
- [3] Access Now Europe and European Center for Not-for-profit Law. Towards Meaningful Fundamental Rights Impact Assessments under the DSA. Retrieved from https://ecnl.org/sites/default/files/2023-09/Towards%20Meaningful%20FRIAs%20under%20the%20DSA_ECNL%20Access%20Now.pdf
- [4] Jack Balkin. 1999. Free Speech and Hostile Environments. *Columbia Law Rev.* 99, 8 (December 1999), 2295–2320.
- [5] Joan Barata. 2024. Threatening dynamics on Internet governance: the case of Europe and the Digital Services Act (DSA). Blog CELE. Retrieved from <https://observatoriolegislativocele.com/threatening-dynamics-on-internet-governance-the-case-of-europe-and-the-digital-services-act-dsa/>
- [6] Joan Barata and Elena Lazar. 2025. Will the DSA Save Democracy? The Test of the Recent Presidential Election in Romania. Tech Policy Press. Retrieved March 24, 2025 from <https://www.techpolicy.press/will-the-dsa-save-democracy-the-test-of-the-recent-presidential-election-in-romania/>
- [7] Anu Bradford. 2023. *Digital Empires: The Global Battle to Regulate Technology*. Oxford University Press.
- [8] Brendan Carr. 2005. Letter to Internet companies. Retrieved from <https://www.fcc.gov/sites/default/files/Chairman-Letter-to-Big-Tech-on-Digital-Services-Act.pdf>.
- [9] Anupam Chander. 2024. When the Digital Services Act goes global. *Berkeley Technol. Law J.* 38, 3 (January 2024), 1067–1088. <https://doi.org/10.15779/Z38RX93F48>
- [10] Agustina Del Campo, Nicolás Zara, and Ramiro Álvarez Ugarte. 2025. Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation (Forthcoming). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5161173
- [11] Agustina Del Campo, Nicolas Zara, and Ramiro Álvarez-Ugarte. 2025. Are Risks the New Rights? The Perils of Risk-based Approaches to Speech Regulation. *J. Intellect. Prop. Inf. Technol. Electron. Commer.* 16, 2 (2025). <https://doi.org/10.2139/ssrn.5161173>
- [12] Evelyn Douek. 2021. The Limits of International Law in Content Moderation. *UC Irvine J. Int. Transnatl. Comp. Law* 6, 1 (May 2021), 37–76.
- [13] ECtHR. 1976. *Handyside v. the United Kingdom*. Retrieved January 28, 2025 from <https://hudoc.echr.coe.int/eng?i=001-57499>
- [14] European Union. Code of Conduct on Disinformation.

- [15] Alexandra Geese. 2023. Why the DSA Could Save Us From the Rise of Authoritarian Regimes. *Van Hoboken J Al “Putting the DSA into Practice. Enforcement, Access to Justice and Global Implications. Verfassungsbooks* (2023).
- [16] Global Network Initiative and Digital Trust & Safety Partnership. 2024. *European Risks and Rights: Stakeholder Engagement Forum. Event Summary*. Retrieved June 3, 2025 from <https://globalnetworkinitiative.org/wp-content/uploads/GNI-DTSP-Forum-Summary.pdf>
- [17] Eric Goldman and Jess Miers. 2021. Regulating Internet Services by Size. Retrieved March 28, 2025 from <https://papers.ssrn.com/abstract=3863015>
- [18] David Kaye. 2018. *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression on the regulation of user-generated online content*. Retrieved from https://digitallibrary.un.org/record/1631686/files/A_HRC_38_35-EN.pdf
- [19] David Kaye. 2024. The Risks of Internet Regulation. *Foreign Affairs*. Retrieved March 25, 2025 from https://www.foreignaffairs.com/united-states/risks-internet-regulation?check_logged_in=1&utm_medium=promo_email&utm_source=lo_flows&utm_campaign=article_link&utm_term=article_email&utm_content=20250325
- [20] Gianclaudio Malgieri and Cristiana Santos. 2025. Assessing the (severity of) impacts on fundamental rights. *Comput. Law Secur. Rev.* 56, (April 2025), 106–113. <https://doi.org/10.1016/j.clsr.2025.106113>
- [21] Martin Husovec. 2022. Will the DSA work? *Verfassungsblog*. Retrieved October 3, 2025 from <https://verfassungsblog.de/dsa-money-effort/>
- [22] Martin Husovec. 2023. Rising Above Liability: The Digital Services Act as a Blueprint for the Second Generation Of Global Internet Rules. *Berkeley Technol. Law J.* 38, 3 (2023), 883–920.
- [23] Martin Husovec. 2024. The Digital Services Act’s red line: what the Commission can and cannot do about disinformation. *J. Media Law* 16, 1 (2024), 47–56.
- [24] Matías González Mama and Ramiro Álvarez Ugarte. 2025. Regulatory Modeling. Tracking the Influence of the Digital Services Act in Latin America. <http://dx.doi.org/10.2139/ssrn.5197380>
- [25] Natalie Alkiviadou. 2025. Why We Should Be Concerned about The “Illegal Hate Speech” Code of Conduct +. *The Bedrock Principle*. Retrieved March 11, 2025 from <https://www.bedrockprinciple.com/p/why-we-should-be-concerned-about>
- [26] Oliver Marsh. *Researching Systemic Risks under the Digital Services Act*. Algorithm Watch. Retrieved from https://algorithmwatch.org/en/wp-content/uploads/2024/08/AlgorithmWatch-Researching-Systemic-Risks-under-the-DSA-240726_v2.pdf
- [27] Paddy Leerssen. 2024. Jawboning and lawboning: comparing platform-state relations in the US and Europe. *Blog CELE*. Retrieved from <https://observatoriolegislativocele.com/jawboning-and-lawboning-comparing-platform-state-relations-in-the-us-and-europe/>
- [28] Paddy Leerssen. 2024. Outside the Black Box: From Algorithmic Transparency to Platform Observability in the Digital Services Act. *Weizenbaum J. Digit. Soc.* 4, 2 (May 2024), 1–29.

- [29] Andrea Palumbo. 2024. A Medley of Public and Private Power in DSA Content Moderation for Harmful but Legal Content: An Account of Transparency, Accountability and Redress Challenges. *JIPITEC – J. Intellect. Prop. Inf. Technol. E-Commer. Law* 15, 3 (December 2024), 246–268.
- [30] Pietro Ortolani. 2023. If You Build It, They Will Come: The DSA’s “Procedure Before Substance” Approach. *Van Hoboken J AI “Putting DSA Pract. Enforc. Access Justice Glob. Implic. Verfassungsbooks* (2023).
- [31] Rachel Griffin. 2023. Codes of Conduct in the Digital Services Act: functions, benefits and concerns. *Technol. Regul.* 2024, (May 2023), 167–187. <https://doi.org/10.2139/ssrn.4463874>
- [32] Rachel Griffin. EU Platform Regulation in the Age of Neo-Illiberalism. <http://dx.doi.org/10.2139/ssrn.4777875>
- [33] Ramiro Álvarez Ugarte. 2024. Bad Cover Versions of Law: Inescapable Challenges and Some Opportunities for Measuring Human Rights of Corporate Conduct in the ICT Sector.
- [34] Rep. Jim Jordan. 2025. Letter from US Representative Jim Jordan to Commissioner Henna Virkkunen. Retrieved from https://www.politico.eu/wp-content/uploads/2025/01/31/document-2025_1_42.pdf
- [35] Reva Siegel. 2004. The Jurisgenerative Role of Social Movements. June 2004. Oaxaca, Mexico. Retrieved from https://law.yale.edu/sites/default/files/documents/pdf/Faculty/Siegel_Jurisgenerative_Role_of_Social_Movements.pdf
- [36] Sally Broughton Micova and Andrea Calef. 2023. Elements for Effective Systemic Risk Assessment Under the DSA. *SSRN Electron. J.* (2023). <https://doi.org/10.2139/ssrn.4512640>
- [37] Barrie Sander. 2020. Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation. *Fordham Int. Law J.* 434, 4 (2020), 939–1006.
- [38] Susan Benesch. 2020. But Facebook’s Not a Country: How to Interpret Human Rights Law for Social Media Companies. *Yale J. Regul. Bull.* 38, (2020), 86–111.
- [39] Théophane Hartmann. 2023. EU’s Breton urges Musk to tackle spread of disinformation on X after Hamas attack. *Euractiv*. Retrieved November 3, 2025 from <https://www.euractiv.com/section/global-europe/news/eus-breton-urges-musk-to-tackle-spread-of-disinformation-on-x-after-hamas-attack/>
- [40] Thierry Breton [@ThierryBreton]. 2024. With great audience comes greater responsibility #DSA As there is a risk of amplification of potentially harmful content in [redacted] in connection with events with major audience around the world, I sent this letter to @elonmusk [redacted] <https://t.co/P1IgxPLzn>. *Twitter*. Retrieved September 5, 2024 from <https://x.com/ThierryBreton/status/1823033048109367549>
- [41] Donald J. Trump. 2025. *Defending American Companies and Innovators From Overseas Extortion and Unfair Fines and Penalties*. Retrieved from <https://www.whitehouse.gov/presidential-actions/2025/02/defending-american-companies-and-innovators-from-overseas-extortion-and-unfair-fines-and-penalties/>
- [42] James David Vance. 2025. Vice President Vance Remarks at French AI Summit. *C-Span*. Retrieved from <https://www.c-span.org/event/public-affairs-event/vice-president-vance-remarks-at-french-ai-summit/430855>
- [43] Carl Vander Maelen and Rachel Griffin. 2023. Twitter’s retreat from the Code of Practice on Disinformation raises a crucial question: are DSA codes of conduct really voluntary? *DSA Observatory*. Retrieved March 11, 2025 from <https://dsa-observatory.eu/2023/06/12/twitters-retreat-from-the-code-of-practice-on-disinformation-raises-a-crucial-question-are-dsa-codes-of-conduct-really-voluntary/>
- [44] Folkert Wilman. 2022. The Digital Services Act (DSA): an Overview. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4304586

- [45] 1964. *Jacobellis v. Ohio*.
- [46] 1979. *The Sunday Times v. the United Kingdom (no.1)*. Retrieved from en: <https://hudoc.echr.coe.int/eng?i=001-57584>,
- [47] 1986. *Lingens v. Austria*. Retrieved May 5, 2025 from <https://hudoc.echr.coe.int/fre?i=001-57523>
- [48] 2007. *Lindon, Otchakovsky-Laurens and July v France*.
- [49] 2011. *General Comment No. 34 on Article 19 of the ICCPR*.
- [50] 2015. *Perinçek v. Switzerland (ECHR)*.
- [51] 2024. *Communication from the Commission – Commission Guidelines for providers of Very Large Online Platforms and Very Large Online Search Engines on the mitigation of systemic risks for electoral processes pursuant to Article 35(3) of Regulation (EU) 2022/2065*. Retrieved March 11, 2025 from <http://data.europa.eu/eli/C/2024/3014/oj/eng>
- [52] *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)*.
- [53] Digital Services Act: Questions and Answers | Shaping Europe’s digital future. Retrieved March 6, 2025 from <https://digital-strategy.ec.europa.eu/en/faqs/digital-services-act-questions-and-answers>
- [54] *European Convention on Human Rights*.
- [55] Commission publishes guidelines under the DSA. *European Commission - European Commission*. Retrieved March 11, 2025 from https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707
- [56] The Code of Conduct on Disinformation | Shaping Europe’s digital future. Retrieved March 11, 2025 from <https://digital-strategy.ec.europa.eu/en/library/code-conduct-disinformation>
- [57] The Code of conduct on countering illegal hate speech online + | Shaping Europe’s digital future. Retrieved March 11, 2025 from <https://digital-strategy.ec.europa.eu/en/library/code-conduct-countering-illegal-hate-speech-online>
- [58] The Commission sends request for information to X under DSA. *European Commission - European Commission*. Retrieved March 6, 2025 from https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4953