**October 21, 2021**

# Joint Statement of the Digital Services Act Human Rights Alliance

## The European Union Should Set a High Bar To Protect Fundamental Rights

The impacts of Internet legislation are rarely contained by borders, and that is particularly true when that legislation comes from influential bodies like the European Union. The "Digital Services Act Human Rights Alliance" (DSA Alliance) was formed partially in response to the ongoing and visible global impact of EU Internet legislation. To date, the EU has been a leader in Internet legislation, for better and, unfortunately, sometimes for worse, and the Digital Services Act is no different.

The DSA Alliance believes that by focusing on the protection of fundamental rights globally, the DSA is an opportunity for the European Union to provide a positive global example in lawmaking. Without such focus on the most pressing challenges for users across the world, there is a risk that the DSA will spawn copycats throughout the world, which could export negative elements to other countries, as we have seen with the German NetzDG. The NetzDG has been criticized even in Germany—for example, a recent study showed the regulation isn't helping remove problematic speech, but is potentially leading to overblocking. Yet countries that offer far less protection for free expression continue to point to NetzDG as inspiration for new laws.

## The DSA's Global Scope

We worry about the standards that the DSA regulation is setting. Although these standards will influence platforms' operations far beyond the Union, no attempt has been made by the EU Commission to consider the needs of—and risks to—vulnerable groups or marginalized communities around the world.

### Protect Access to Information Rights: Legal Representatives

***We urge EU co-legislators to avoid disproportionate demands on smaller providers that would put users' access to information in serious jeopardy.***

An example of a problematic standard are the rules for appointing legal representatives in the European Union. Due to the wide territorial scope of the DSA, non-EU platforms that provide services to the EU, including small and micro providers, will have to appoint a costly legal representative, who must be based in the EU (Article 11). There is thus a huge risk that platforms, in particular smaller ones, could stop offering services to the EU altogether. For example, some news sites from the US stopped offering access to EU users when new GDPR regulations went into effect in 2018. The risk that users could lack access to information was recognized in the EU Commission's Impact Assessment to the DSA, which explained that an exception should be made for "very small providers." However, such an exception was not included in the DSA.

Another example are the provisions on trusted flaggers, which are problematic for several reasons already in the European context (as elaborated on in the section below). Outside of the EU, the formalization of trusted flagger relationships has proved very dangerous, such as in Myanmar. Platform decisions tend to be perceived as the result of influence by local actors, rather than a platform's own decision, and flaggers, rather than platforms, are blamed for unsatisfactory decisions.

## Content removals

### Short and Inflexible Deadlines Conflict With Fundamental Rights

***EU co-legislators should avoid legally mandated strict and short time frames for content removals due to their detrimental impact on the right to freedom of expression and opinion.***

We are particularly worried about the inclusion of unduly short time frames and a strong push for swift content removals that are prominent in recent opinions on the DSA proposal issued by parliamentary committees in the European Parliament.

First, short time frames make it impossible for platforms to carry out a "legality check," let alone make a decision about whether to act against content that requires contextual analysis. Due to their inflexible nature, strict deadlines also make it difficult for providers to prioritize decisions and to act according to their resources. Small providers will likely not have a system in place that allows them to react that quickly. Even larger service providers may not be able to deal with a high number of removal requests.

For this reason, the EU's e-Commerce Directive has abstained from introducing a fixed period to act after a service provider has become aware or obtains knowledge of illegal activity. Instead, the Directive explains that the removal or disabling of access has to be undertaken "expeditiously," in "observance of the principle of freedom of expression." We understand the need to establish standards for platforms to react to notices and follow up to avoid harm to users. We also support increased responsibility for platforms to enforce their <u>own</u> standards and terms of services. However, we urge the European co-legislators not to abandon the fundamental rights perspective by imposing inflexible time frames for removals of third party content that inevitably will lead to removal of lawful content to avoid liability. Under human rights principles, any restriction on freedom of expression must be necessary and proportionate: such measures should not go beyond what is necessary to achieve the objectives. Short deadlines for removal of alleged illegal content fail to balance takedown objectives with users' right to free speech online, as demonstrated by the French Online Hate Speech Bill. Because of its negative impact on users' right to freedom of expression, the bill was declared unconstitutional by the Constitutional Council of France.

## Algorithmic Content Moderation: Recognize Limits and Protect Fundamental Rights

***We urge EU co-legislators not to impose legally mandated automated content moderation tools on online platforms, as this will lead to over-removals of legitimate speech.***

In order to meet such short deadlines in a cost-effective way, online platforms mostly rely on faulty algorithmic content moderation tools. These tools not only have limited ability to assess context, in their current form their decision-making processes are completely opaque. Legislators often sidetrack the proper safeguards and legally mandated requirements for meaningful transparency. However, without these safeguards, users cannot have access to appeal mechanisms, effective remedy and other safeguards of procedural fairness. These should be an integral part of any platform governance legislation.

Increased pressure on platforms to remove more content faster can also result in general monitoring of user content, banned under EU rules. Unduly short deadlines to assess the legality of content and disproportionate obligations to comply with removal requests without the option of derogation or extension will often leave platforms no other choice but to resort to automated content filtering systems. However, filters are notoriously inaccurate, prone to overblocking legitimate material, and lacking in checks and balances.

Experience with the already increased use of algorithmic moderation to detect content related to terrorism has demonstrated this issue—huge amounts of content documenting human rights abuses have already been removed through automated moderation, with little accountability. In fact, the issue is problematic enough that the Facebook Oversight Board recently advised Facebook to increase transparency around its existing use of automation. Adding even more algorithmic content moderation now, when these issues haven't been resolved, is particularly dangerous.

This concern is amplified by frequent demands for platforms to detect and restrict content that has been previously identified as illegal, or is similar to such illegal content. We support a dialogue about how platforms can better address concerns about illegal content being multiplied in bad faith. However, the DSA should acknowledge that each time a post is being shared and re-shared by online users, its context and motivation may be drastically different. The list of possible legitimate re-uploads is long: from journalism to satire, humor, and academic purposes. Even if the initial post was deemed illegal, each re-upload may change it in a materially different way. Algorithmic content moderation tools are currently unable to assess these contexts and will inevitably fail to grasp nuance, no matter how much the tools improve.

The proposed DSA is an opportunity to determine how online platforms safeguard people's fundamental rights, such as privacy and freedom of expression. Policy makers should concentrate on helping users to better understand how moderation decisions are made, hence regulating "process" rather than speech. If strong human rights and due diligence safeguards are not sufficiently incorporated in content governance legislation, it will become an empty shell fuelling illegitimate

restrictions and digital suppression of freedom of expression under the disguise of protecting society against illegal content online.

### Quick Fixes Don't Work in Practice

***EU co-legislators should avoid shifting states' obligations to protect individuals' rights to privately-owned online platforms, thus allowing them to act as quasi-judicial bodies in the online ecosystem without any public scrutiny.***

Furthermore, shortsighted and quick fixes have already proved to miss their mark in practice. In order to comply with regulatory demands, online platforms will assess the content against their terms of service and quickly remove it outside of any public scrutiny. For instance,to comply with the German Network Enforcement Act (NetzDG) that 'promoted' swift content removals under short time frames as a silver bullet solution to illegal content online, Facebook created a two-tiered procedure. First, all reported content is reviewed under its Community Standards. If a violation is found, the piece of content will be removed globally. It will never be judged against the NetzDG provisions, nor will it be included in the mandatory transparency report. Second, if content violates NetzDG but it is in conformity with Facebook's own rules, it will be blocked only in Germany. This has led to NetzDG practically becoming "the community standards enforcement law".

## Enforcement Overreach and Who Should be Trusted

### When Information is Handed to Law Enforcement Authorities: Protect User Privacy

***EU co-legislators should not impose mandatory reporting obligations to Law Enforcement Agencies (LEAs), especially without appropriate safeguards and transparency requirements.***

The DSA contains provisions that demand cooperation with law enforcement authorities. For example, online platforms must inform enforcement authorities about serious criminal offences. It is troubling that this notification duty is not restricted to imminent risks and does not specify which data platforms must share. Suggestions within the responsible committees in the EU Parliament are even more concerning, such as mandatory data retention and transmission of content to authorities. Such rules would undermine the privacy of users. And they are likely to have disparate impact on people who are already suffering from discrimination based on race or religion, and will have a predictable chilling effect on freedom of expression, as users cannot communicate freely if they are concerned that uploaded content is shared with law enforcement authorities.

### Trusted Flaggers Should be Trustworthy and Independent of Law Enforcement and Companies

***We urge European co-legislators to prevent public authorities, including LEAs, from becoming trusted flaggers. Conditions for becoming trusted flaggers need to be subject to regular reviews and proper public oversight.***

We are also concerned about the provision on trusted flaggers (Article 19, Recital 46), whose notices must be processed and decided upon with priority. The DSA proposal rightly sets criteria for becoming a trusted flagger, but falls short of other necessary safeguards. Any legal framework establishing a model of platform governance should contain proper legal safeguards for the impartiality and independence of those trusted flaggers. Trusted flaggers are entities with specific expertise and dedicated structures for detecting and identifying illegal online content, independent of both government and private companies. However, under the DSA, law enforcement agencies and industry-oriented entities can also be trusted flaggers (e.g. in the realm of copyright). European co-legislators ought to ensure that under no circumstances should the conditions for instituting trusted flaggers be determined solely by private platforms or allow public authorities to assume this role.

It is particularly worrisome that under current wording of Article 19 law enforcement agencies may become trusted flaggers. The weakness of this provision should be viewed against a number of proposals currently being discussed in the European Parliament to extend data access and reporting obligations by platforms to law enforcement agencies. The lack of aforementioned safeguards could impose a significant risk of human rights abuse, especially in those countries that have been experiencing democratic backsliding and ongoing weakening of the rule of law. The EU lawmakers should also bear in mind that many companies already have designated contacts and processes for removing content and sharing information in cooperation with law enforcement, often based on existing policies such as the EU Crisis Response Protocol. Weaving law enforcement agencies into the trusted flagger provision is not only unnecessary, it threatens to undermine due process.

## Notice-and-Action System and Knowledge-Based Liability

Online platforms are vital to the vast majority of internet users around the world. The policies intermediaries adopt shape users' social, economic, and political lives. They also have major implications for users' fundamental rights, including freedom of expression, freedom of association, and the right to privacy. We are concerned at the growing number of governments around the world that are embracing heavy-handed approaches to intermediary regulation, and believe that the current proposal could provide repressive regimes with cover for these approaches, as NetzDG has done.

### Preserve the "Actual Knowledge" Concept

***We urge EU co-legislators to preserve the current conditional model of intermediary liability, as established by the E-Commerce Directive.***

It is very concerning that under the proposed notice-and-action mechanism, as presented by the EU Commission, properly substantiated notices automatically inform platforms of the notified content. As host providers only benefit from limited liability for content users post and share when they expeditiously remove content they "know" to be illegal, there is a risk that platforms will have no other choice than to over-block notified content to escape the liability threat and use upload filters, as suggested by the DSA proposal (Article 14(6)).

The latest opinions of responsible parliamentary committees go a step further and suggest that hosting providers should not benefit from liability exemption if they do not comply with newly established due diligence obligations in the DSA proposal. Some proposals amending the current DSA draft even suggest that all intermediaries that assist users in optimising, classifying, organising, referencing or promoting content, or have control over it should be excluded from liability exemption if they do not comply with due diligence safeguards. Other proposals suggest subjecting online platforms to potential liability for user content if they fail to comply with mere ancillary duties, such as being transparent in terms of service or providing information about consumer products. We sympathize with public policy concerns related to the increasingly powerful role intermediaries play in modern society and the need to ensure that they act responsibly. However, placing the knowledge-based intermediary liability regime under direct threat would establish a dangerous precedent across the world. Instead of modifying current liability rules and making online platforms the internet police, EU policymakers should focus on rules that empower users, protect marginalized groups, and provide procedural justice rights in case things go wrong. In this respect, the EU Commission's DSA proposal got many things right, as it set out a new transparency standard content removals and other due diligence obligations focusing on platforms' operations, such as criteria for platform accountability and independent auditing.

## Risk Assessment and Mitigation of Risks Measures

The DSA establishes a new set of due diligence obligations for very large online platforms (VLOPs). VLOPs must conduct assessments of significant systemic risks stemming from the functioning and use of their services in the EU. Systemic risks include the dissemination of illegal content and negative effects on fundamental rights. Platforms must also put in place reasonable, proportionate, and effective risk mitigation measures. There is no requirement in the text that such measures must be "necessary" and there are no standards in place for such risk assessments. It's up to platforms to decide which mitigation measures to take. The relevant provisions do not require a human rights impact assessment and encourage the set-up of codes of conduct. The role of civil society in all of this is unclear.

## Strengthen Mandatory Human Rights Impact Assessments

***EU co-legislators should consider mandatory human rights impact assessment as the primary mechanism for examining and mitigating systemic risks stemming from platforms' operations. Mitigation of risk measures should play a secondary role only.***

Mandatory human rights impact assessments (HRIA) should be strengthened in the DSA proposal, especially when it comes to identifying and assessing systemic risks in VLOP functions and use of their services. Currently, the DSA only requires an HRIA once a risk has been identified, which would greatly limit their scope. We are convinced that the assessment of all VLOP operations, including use of algorithmic decision-making, should be conducted to analyze impacts on human rights and not be limited to a risk mitigation exercise. The burden of proof should be on VLOPs to demonstrate that their services as a whole and their individual products and technical tools do not violate human rights.

We would also like to stress that "rights granted to the data subject by EU law should be respected regardless of the level of the risks which the latter incur through the data processing involved." We must therefore avoid a situation in which VLOPs can shirk their responsibilities by ignoring user rights. The legislators should clarify that a risk assessment should be publicly available and complement, not replace, obligations to respect fundamental rights.

We support the DSA's approach to subject VLOPs to independent auditing, which should also cover risk assessments. Third-party independent auditing of algorithmic systems should be mandatory, particularly on those online platforms that deploy algorithmic tools for content moderation and curation, such as content recommender systems, advertising, and microtargeting.

## Guidance on How to Design HRIA

Finally, European co-legislators and policy makers should focus in the future on specifying the contents of mandatory human rights impact assessments. . The DSA Human Rights Alliance suggests that the following questions be considered:

**WHAT?** Should HRIAs apply to a specific product or a type of technology? Should HRIAs be limited to some AI systems only, or expand to all?

**WHERE?** Should a report focus on one specific jurisdiction (e.g. European Union) or take a global approach across multiple jurisdictions (e.g. UN, Council of Europe, country or regional-level analyses)?

**WHEN?** At what stage should HRIAs be conducted (ex ante; ex post; ongoing) so that the impacts are best captured and specific context is considered?

**WHO?** What kinds of skills and expertise are needed and, if a regulatory agency is responsible for conducting the HRIA, what access would it need, and what kind of third party would be deputized to do it?

**HOW?** How do we want to assess risk? Because risk levels are subjective, should we instead focus on risk indicators/signals, which would trigger a necessary regulatory response?

**IMPACT?** Would HRIAs actually prevent harm, or would they promote uptake of AI systems? Is the scope of HRIAs broad enough to include all types of harm related to algorithmic systems? How can we prevent "participation washing" of civil society engagement in the process of conducting HRIAs for AI? How can we ensure that the findings of HRIAs are integrated and acted on?

**ORGANIZATIONS**

7amleh - The Arab Center for the Advancement of Social Media
Access Now
Center for Democracy and Technology
Civil Liberties Union for Europe (Liberties)
Electronic Frontier Foundation (EFF)
European Center for Not-for-Profit Law Stichting (ECNL)
Global Forum for Media Development (GFMD)
Global Voices
Mnemonic (Syrian Archive)
Open Technology Institute
R3D: Red en Defensa de los Derechos Digitales (R3D)
Ranking Digital Rights