bff Platform monitoring with a focus on Article 12, 16, 34 and 35 of the DSA

Platform regulation must prioritise protection against violence

A policy paper by the bff

Bundesverband
Frauenberatungsstellen
und Frauennotrufe

bff Platform monitoring with a focus on Article 12, 16, 34 and 35 of the DSA

Platform regulation must prioritise protection against violence

The *Digital Services Act (DSA)* is the central European law regulating large online platforms. Its aim is to make digital spaces safer, protect fundamental rights and reduce systemic risks such as hate, disinformation and gender-based violence. For victims of *(digital) gender-based violence* – such as non-consensual intimate images (NCII), deepfakes, stalking or doxing – a single click can have enormous, sometimes life-changing consequences. Platforms play a central role in this: they decide what is visible, how content is distributed and how survivors are supported.

For the first time, the DSA obliges very large online platforms (*VLOPs*) and search engines (*VLOSEs*) to **actively identify and mitigate systemic risks**, provide **user-friendly reporting mechanisms** and set up **single points of contact** for recipients of their services. However, our monitoring shows that despite a strong legal basis, implementation often fails in practice – with serious consequences for the survivors.

This policy paper shows

- **which obligations are particularly relevant for survivors of gender-based violence,**
- **where platforms are failing**,
- and **which feminist, practice-oriented policy measures are now necessary** to ensure protection against (digital) gender-based violence.

Our analysis is based on the systemic risk reports pursuant to Art. 34 DSA from 2023, which we evaluated from the major social platforms (Snapchat, TikTok, Instagram, Facebook) and porn platforms (XVideos, Pornhub and Stripchat).

Important articles at a glance

Art. 34 DSA – Risk assessment	VLOPs and VLOSEs must regularly identify and assess systemic risks, including explicitly risks of gender-based violence.
Art. 35 DSA – Mitigation of risks	Based on the risk analyses, effective measures must be implemented and their effectiveness continuously reviewed (e.g. reporting mechanisms, moderation, algorithmic adjustments, cooperation with experts).
Art. 12 DSA – Single points of contact (SPOC)	Platforms must set up a central, easily accessible single point of contact (SPOC) through which survivors and counselling centers can communicate directly with the platform.
Art. 16 DSA – Easily accessible reporting channels	Platforms are obliged to provide user-friendly and easily accessible reporting mechanisms.

WHERE DOES IT FAIL IN PRACTICE?

Deficits in Art. 34/35 – Risk assessments & risk reduction

1. **Narrow or missing definitions of gender-based violence**
   Many platforms define gender-based violence too narrowly and usually without an intersectional perspective. Discrimination based on gender, race, class, disability, sexuality or migration is rarely considered together – even though it is precisely these intersections that determine how vulnerable people are to gender-based violence (online). Furthermore, many platforms do not recognise key forms of gender-based digital violence, such as *doxing* or *sexualised deepfakes*, as such. These attacks are often perpetrated within close social circles – for example, by (ex-)partners, family members or people from the victim's social circle – but are not classified as gender-based violence in risk reports.

2. **Lack of connection to real-world contexts of violence**
   Gender-based violence is viewed as a purely digital phenomenon. Digital violence as a continuation of (ex-) partner violence, psychological violence, control and surveillance in the analogue world is not considered by platforms, even though these forms are closely intertwined.

3. **Superficial risk assessments**
   In order to better protect users from illegal content, many platforms have named measures such as blocking, reporting mechanisms or AI filters, but their effectiveness and target group relevance remain unclear. Business models that generate risks through advertising and engagement optimisation are hardly reflected upon. Many reports seem like pure 'show compliance' without in-depth assessment.

4. **Lack of transparency**
   Reports are divided/subdivided into categories, time periods and formats in an inconsistent manner, are often difficult to find or are even redacted in places (e.g. TikTok). Contrary to studies, Snapchat classifies sexualised violence as an 'extremely low risk'. The lack of standards makes comparability and external control difficult.[1]

5. **Advertising as an underestimated risk**
   For months, deepfake/*nudify* apps were advertised via Meta Ads [2] – in some cases also visible to underage users who are hardly able to assess the implications of such tools. At the same time, Google Ads continues to display advertisements for *stalkerware*[3]. Advertising is hardly recognised by platforms as an independent risk for gender-based violence, even though it promotes violence.

1 Snapchat, The Digital Well-Being Index – Key Research Results, February 2023: https://assets.ctfassets.net/kw9k15zxztrs/f8tyHpE9HzLT9bGpPIg9B/e92cf6506b284bf4d27f0fef6bb1e264/DWBI_Findings_English.pdf
2 Quelle: 404 Media, 15.01.2025, „Instagram Ads send this nudify Site 90 Percent of its traffic", https://www.404media.co/instagram-ads-send-this-nudify-site-90-percent-of-its-traffic/
3 Quelle; August 2024, Complaint against Google Ireland Limited regarding Art. 35 (1) in conjunction with Art. 34 (1) DSA, submitted by Gesellschaft für Freiheitsrechte e.V. in cooperation with the project "Ein Team gegen digitale Gewalt", https://freiheitsrechte.org/uploads/documents/Englische-Dokumente/Freedom-in-the-digital-Age/Complaint-Google-Stalking-Apps.pdf

Recommendations on Articles 34/35 DSA

- Harmonisation of risk assessments through minimum standards for categories, time periods and methodology.
- Establishment of a permanent, public and machine-readable archive for all DSA documents to improve transparency and comparability.
- Platforms must recognise gender-based violence as a human rights violation and make this visible in their risk analyses.
- Development and binding implementation of a common, intersectional understanding of violence based on the Istanbul Convention.
- Conducting holistic risk assessments that combine technical, social and structural factors of violence.
- Gender-based violence is not an individual risk, but has a systemic effect across moderation, advertising, algorithmic amplification and design decisions. Risk analyses must mandatorily reflect this cross-cutting nature.
- Risk mitigation measures must be explicitly gender-sensitive and intersectional in order to recognise increased risks for different affected groups.
- Platforms must systematically involve civil society organisations and survivors in the preparation of their risk assessments and remunerate them for their participation.
- Introduction of regulated standards for the effectiveness, target group orientation and evaluation of risk mitigation measures.
- Platforms must disclose what data they collect and use, and systematically evaluate and publish measures according to target groups, effectiveness and the specific vulnerability of different groups.
- Clear recognition of sexualised deepfakes as gender-based violence.
- Inclusion of the promotion and distribution of programmes that enable non-consensual sexualised deepfakes as an independent risk factor for gender-based violence.
- Platforms are already obliged to actively mitigate risks; however, this obligation must be consistently implemented, in particular through the systematic removal of advertisements for such applications.

Deficits in Art. 12 – Single Points of Contact (SPOCs)

1. **Contact points that are difficult to find:**
   SPOCs are often hidden deep in the menu and difficult to access, requiring several clicks.
2. **Manipulative interface design:**
   Endless scrolling or small footers make the contact point practically invisible.
3. **Language barriers**
   Many SPOCs are only available in English or in legally complicated language.
4. **Lack of focus on data subjects**
   Instead of specific contact persons, data subjects often only find unclear forms or FAQ pages.
5. **Inconsistent and confusing forms**
   Many platforms lack a transparent explanation of what a SPOC or DSA complaint is and what rights data subjects have.

Recommendations for Art. 12 DSA

- Place SPOCs clearly and visibly in the header or main menu.
- Ensure direct, tamper-free access without hidden click paths.
- Offer simple, accessible language and automatic language selection.
- Integrate a reference to specialised help (e.g. in the European context this would be the referral to the EU-wide helpline 116 016).
- Provide a transparent explanation of the function of SPOCs and the rights of users under the DSA (reporting, complaints, access to *trusted flaggers / out-of-court dispute settlement bodies*).

Deficits in Art. 16 – Reporting mechanisms

The most serious problems lie at the heart of the DSA: **the reporting mechanisms themselves.**
Quote from peer expert Patricia Gutsche from the bff:

> **'I feel fundamentally empowered when I can report something online. But the reporting channels need to be much simpler so that I can actually use them.'**

The study published in October 2025 by Das NETTZ on the implementation of DSA reporting mechanisms on very large platforms is one of the first DSA studies to show the gap between legal requirements and practical usability. The study shows that the DSA reporting channels provided by the EU are hardly used by many users because they are difficult to find and unclear in design. Instead, those affected predominantly report violations of the terms and conditions via the conventional reporting mechanisms, which appear simpler but do not offer the same legal protection. Around a quarter of DSA reports are abandoned prematurely – an indication of significant barriers to use. Many affected parties are also unaware that the DSA grants them special rights, which means that content is often not reported through the correct mechanisms. The result is a structural underutilisation of the DSA's central protection mechanisms.[1]

1. Unclear categories

Those affected must choose between criminal law terms that they do not understand. On Snapchat, it is not clear to users whether they are submitting a report under the DSA or for a violation of the terms and conditions.

2. *Deceptive Patterns*

Platforms deliberately steer affected parties away from DSA reporting mechanisms and towards weaker terms and conditions reports. The reporting mechanisms themselves contain deceptive patterns, such as misleading categories that appear to be legal options but do not trigger a DSA complaint, or extended click paths with repeated confirmations that lead to '*click fatigue*'.

3. High barriers to access

Requirements such as account creation, mandatory address or proof of identity exclude vulnerable groups, e.g. people with uncertain residence status.
In addition, many of the reporting forms give the impression of making an official report or of user liability for false reporting.

4. Contact

Reports often end up directly in automated systems. AI is hardly able to reliably recognize *NCII*, deepfakes, stalking and dependency relationships – even for experts, such cases are complex. Furthermore, it is unclear what data the models are trained with. TikTok, for example, seems to rely on automated moderation without making this fact

1 Quelle: Das NETTZ, Studie: „Zwischen Klick und Konsequenz: Eine Evaluation der Meldeverfahren nach dem Digital Services Act", Oktober 2025, https://www.das-nettz.de/neue-studie-von-das-nettz-zeigt-durch-den-dsa-vorgegebene-meldeverfahren-auf-grossen-online

transparent to users. Such unlabelled automated decisions are a clear violation of Articles 16 and 20 of the DSA.[2]
This means that survivors of gender-based violence are deprived of human, context-sensitive assessments – often exactly what they urgently need.

5. Revictimising processes

A lack of explanations, algorithmic misjudgements and automated rejections reinforce feelings of shame, powerlessness and despair among survivors.

6.Lack of linguistic accessibility

Young people, people with learning difficulties and non-native speakers are particularly at risk of not even finding the reporting channels, or these channels are not usable for these groups.

2 Quelle: 2025, HateAid, „Recht ohne Reichweite – Der DSA im Praxistest“, https://www.stiftung-mercator.de/de/publikationen/hateaid-abschlussbericht-recht-ohne-reichweite/

bff recommendations on Art. 16 DSA – Feminist design of reporting channels

These recommendations were developed collaboratively in a workshop led by the bff titled Reporting Futures at the Digital Futures Gathering on 1–2 October 2025 in Berlin, together with international experts in digital policy and survivor counselling.

Facilitate access

- Define binding EU guidelines on the meaning of easily accessible and user-friendly within the meaning of Art. 16 DSA – including clear minimum standards for language, navigation, accessibility and transparent explanations of terms.
- Uniform, visible reporting button on all platforms.
- 2–3 clearly understandable categories (e.g. image-based violence, stalking, doxing).
- Simple, accessible language with symbols and audio function.
- No risky mandatory fields (address, ID number).
- Multiple access routes and information about them (web, chat, telephone, trusted flaggers).

Ensure protection

- Trauma-informed confirmations and explanations of the process that do not retraumatise survivors.
- Adequate equipment for national *DSCs* and awareness campaigns by them
- Human review instead of purely automated moderation.
- Immediate measures for high-risk content.
- Data protection guarantees.
- Sanctions for manipulative patterns or systematically flawed reporting mechanisms by the EU Commission and national authorities.

Involve survivors

- Co-design of reporting channels with survivors and experts on violence and digital accessibility.
- Trained human content moderators (fairly remunerated, access to supervision, training by violence protection experts).
- A European reporting centre for NCII/ deepfakes that is networked with existing national reporting centres, platforms and authorities and can process reports across systems and platforms (interoperability).
- Transparent key indicators from the perspective of survivors (e.g. average response time, reports that were incorrectly rejected, number of clicks users need to get to the reporting form, abandonment rate, etc.).

Conclusion

The present analyses make it clear: Gender-based (digital) violence is not a marginal phenomenon, but a structural problem that is not adequately addressed by the existing reporting and complaint mechanisms of the platforms. As long as these systems are designed primarily from the perspective of technical efficiency and profit-oriented platform logic, the experiences of those affected – especially people and groups who face multiple discrimination – will remain invisible. A feminist approach therefore means more than just 'better moderation': it requires platforms to acknowledge their responsibility, take intersectional risks seriously and design reporting mechanisms in such a way that they create protection, agency and justice. This includes not ignoring violence in the immediate social environment, but considering it as a central part of gender-based violence. If platforms in the European Union want to create safe digital spaces, there is no way around a feminist, intersectional and survivor-oriented design. These recommendations show what this can look like – and where platforms urgently need to improve.

Contact:

Elizabeth Ávila González
digitalegewalt@bv-bff.de

Glossary

Out-of-court dispute settlement bodies

If a platform deletes a post or account or ignores a complaint, users can contact such a body. It independently checks whether the platform's decision was lawful – without having to go directly to court.

Click fatigue

Click fatigue describes the exhaustion or numbness experienced by users when platforms confront them with too many notifications, selection windows, consent requests or security warnings. Overwhelmed, people often continue clicking in annoyance – which undermines the effectiveness of security or protection mechanisms and can encourage manipulative design.

Deceptive Design (manipulative interface design)

Deceptive design (dark patterns) refers to design tricks in digital interfaces that deliberately push users into making decisions they would not otherwise make – for example, through misdirection, manipulation or the concealment of important options. The aim is usually to collect data, force consent or achieve economic advantages for the platform.

Deepfake

Deepfakes are fake digital representations, mostly photos or videos, created using software. Faces or voices of people are inserted into existing recordings to give the impression that they are saying or doing something that is not actually the case. This technology is often used to create fake pornographic content.

Digital gender-based violence

Digital violence is an umbrella term. It refers to acts of violence committed with the aid of information technology systems:

- violence with technical devices (e.g. smartphones, location trackers, cameras)
- with software (apps, internet applications, emails, etc.)
- and violence in the digital space, e.g. on online portals or social platforms.

It is important to understand digital violence as a continuum of analogue violence: digital means expand and reinforce existing patterns of control, surveillance, intimidation or humiliation. The violence does not change its core – it merely becomes digitalised.

Digital Services Act (DSA)

This is an EU law that obliges large platforms such as Facebook or TikTok to take better action against (digital) violence or disinformation. For example, they must explain how their algorithms work and respond quickly when users report illegal content.

Digital Services Coordinator (DSC)

This is the national supervisory authority for the DSA – in Germany, this is the Federal Network Agency (Bundesnetzagentur). It ensures that platforms comply with the rules and accepts complaints accordingly. It can impose fines or cooperate with other EU bodies.

Doxing
Doxing is the internet-based collection of personal data and the subsequent publication of this data with the aim of exposing and intimidating the person concerned.

Engagement
Engagement describes how strongly users interact with content on a platform. This includes, for example, likes, comments, shares, saves or clicks. Platforms evaluate engagement to decide which content should be made particularly visible. Content with a high level of engagement is often shared more widely – even if it is problematic or violent.

Intersectionality (according to Kimberlé Crenshaw)
Intersectional approaches and analyses take into account that different social categories are conditioned by power relations and, through their interaction and interrelationships, influence (structural) inequalities. In terms of (digital) gender-based violence, this means, for example, that women are attacked online because they are women, but also because they are Black women, or trans, or disabled, or of colour. So it is not just about sexism and gender-based violence, but also about other forms of discrimination, such as racism and ableism. The levels are intertwined and cannot necessarily be separated, but they contribute to different experiences of discrimination.

Non-consensual intimate image (NCII)
Non-Consensual Intimate Images (NCII) refers to the distribution, sharing or publication of intimate images or videos of a person without their explicit consent. This also includes threatening to publish such images. NCII is a form of digitalised gender-based violence and can have serious personal, social and professional consequences for survivors.

Nudify-Apps
Nudify apps are apps or programmes that use artificial intelligence to alter images or videos of people to make them appear naked. The individuals affected have usually not consented to this. The violence perpetrated in this way is predominantly directed at female-presenting individuals in a sexualised manner.

SPOC
Single Point of Contact (SPOC) according to Article 12 DSA is a central, easily accessible contact point for a platform through which users can get in touch or ask questions. The SPOC is intended to ensure that platforms are accessible, especially in cases of illegal or violent content.

Stalkerware
Stalkerware refers to spy apps or programmes that are secretly installed on devices to spy on a person – for example, their messages, locations, photos or online activities. The use of stalkerware is a form of digital violence and is often used in contexts of intimate partner violence to exercise control and surveillance.

Risk Reports
Risk assessments under Art. 34 DSA are mandatory risk analyses that very large online platforms and search engines must regularly carry out and publish. In these reports, they must identify systemic risks – such as the promotion of gender-based violence, discrimination or the dissemination of illegal content – and explain how these risks arise and what impact they have. The reports form the basis for developing appropriate protection and mitigation measures.

Trusted Flagger
These are organisations or bodies that platforms particularly trust. They can report content that violates the law more quickly and with higher priority. Their reports must be given preferential treatment by the platforms.

VLOPs/VLOSEs
These are very large online platforms (VLOPs) or very large online search engines (VLOSEs), such as Instagram, TikTok or Google. They have a particularly large number of active users in the EU – at least 45 million. They are therefore subject to stricter rules under the Digital Services Act. A list of VLOPs/VLOSes can be found **HERE**.

Impressum

Projekt „Aktiv gegen digitale Gewalt“
des Bundesverband Frauenberatungs-
stellen und Frauennotrufe (bff), Berlin

digitalegewalt@bv-bff.de

www.aktiv-gegen-digitale-gewalt.de

bff:
Bundesverband
Frauenberatungsstellen
und Frauennotrufe