Online ISSN 1759-2917
Print ISSN 1759-2909

Working Paper

ODI Europe

Gender equality and digital platforms

What role for the EU's Digital Services Act?

Aline Burni and Divij Joshi

November 2026

Key messages

Online gendered harms and backlash against gender equality encompass gender-based violence, the promotion of 'traditional' gender norms, the suppression of gender equality voices and gendered disinformation. These not only harm individuals; they also undermine democracy and progress towards gender equality, forming part of an increasingly organised and interconnected web of actors, narratives and practices.

While digital platforms do not create gendered harms and gender backlash, they do shape their dissemination, amplification and proliferation. Through features such as recommendation systems, virality mechanisms, content moderation frameworks and cross-platform dynamics, platforms can increase the visibility and reach of harmful content.

The Digital Services Act (DSA) is the EU's strongest framework to date for increasing platform accountability, transparency and oversight, creating important tools to address online gendered harms and gender backlash. Beyond requiring the removal of illegal content, it offers mechanisms to address the systemic risks platforms pose to gender equality, including gender-based violence, the amplification of harmful content and gendered disinformation.

For the DSA to be effective, enforcement must be robust. Success will depend not only on platforms' compliance, but also on the effectiveness of wider co-regulation. This requires adequate resourcing and the integration of gender expertise across the board, as many forms of gendered harms remain difficult to address in practice. Effective coordination between EU and Member State authorities will be critical.

ODI Europe
35 Square de Meeûs
1000 Brussels
Belgium

EU transparency register: 051479091677-17

© ODI Global 2026

Online ISSN 0140-8682

This work is licensed under CC BY-NC-ND 4.0.

Readers are encouraged to reproduce material for their own publications, as long as they are not being sold commercially. ODI Global requests due acknowledgement and a copy of the publication. For online use, we ask readers to link to the original resource on the ODI Global website.

Views and opinions expressed in this publication are the responsibility of the author(s) and should in no way be attributed to the institutions to which they are affiliated or to ODI Global.

How to cite: Burni, A. and Joshi, D. (2026) *Gender equality and digital platforms: What role for the EU's Digital Services Act?* ODI Europe working paper. Brussels: ODI Europe (https://odi.org/en/about/our-work/odi-europe)

Acknowledgements

The authors are grateful to the many experts and practitioners who shared their knowledge and perspectives with us. With their permission, we acknowledge the contributions of Prof. Dr. Jan Tobias Mühlberg, Prof. Dr. Julia Hörnle, Prof. Dr. Kim Barker, Asha Allen, Dr. Silvia Semenzin and Tsisi Matekaire; others preferred to remain anonymous. We would like to thank Claire Kumar, Dora Meredith, Rachel Marcus, Julie Luebken and Dr. Naomi Appelman for their peer review and helpful feedback. We thank Anushka Sisodia for her research assistance in project design and interviews. Thanks to Sherry Dixon for coordinating production, Matthew Foley for editing and Aaron Griffiths for typesetting. Opinions, any errors or omissions remain the responsibility of the authors.

About the authors

Aline Burni is a Research Fellow on Gender and Foreign Policy with ODI Europe.

Divij Joshi is a Research Fellow in the Power and Rights research group at ODI Global.

Contents

Acknowledgements / i

Display items / iii

Abbreviations and acronyms / iv

Executive summary / 1

1 **Introduction** / 5

2 **Digital platforms** / 7
- 2.1 Three types of platforms / 7
- 2.2 Multiple dimensions of platforms / 8

3 **Mapping gendered harms and backlash against gender equality** / 12
- 3.1 Gender-based hate and violence / 14
- 3.2 Reinforcement of 'traditional' gender norms / 17
- 3.3 Suppression of gender equality content, voices and debate / 18
- 3.4 Gendered disinformation / 21
- 3.5 Implications / 22

4 **How platforms shape online gendered harms and gender backlash** / 24
- 4.1 Algorithmic systems and virality / 24
- 4.2 Modularity / 26
- 4.3 Anonymity / 27
- 4.4 Moderation policies and enforcement / 27
- 4.5 Cross-platform dynamics / 29
- 4.6 Interaction of technical and organisational factors / 29

5 **Platform regulation: potential and limitations of the Digital Services Act** / 31
- 5.1 Main DSA requirements / 31
- 5.2 Relevance to gendered harms and gender backlash / 33
- 5.3 Limitations of the DSA in addressing gendered harms and gender backlash / 36

6 **Policy priorities and recommendations** / 40
- 6.1 Strengthen DSA enforcement / 40
- 6.2 Resource civil society to support enforcement / 41
- 6.3 Resist de-regulatory pressure / 41
- 6.4 Emphasise preventative measures and empower users / 42
- 6.5 Integrate platform regulation aspects in international partnerships / 42

References / 44

Appendix 1 Key Informant Interviews / 53

Display items

Boxes

Box 1 Algorithmic systems / 7
Box 2 Case studies / 12
Box 3 Online gender backlash / 13
Box 4 Technology-facilitated gender-based violence / 14

Abbreviations and acronyms

AI	Artificial Intelligence
DSA	Digital Services Act
ECD	e-Commerce Directive
ECJ	European Court of Justice
EU	European Union
IBSA	Image-based sexual abuse
Incels	Involuntary celibates
LGBTQIA+	Lesbian, Gay, Bisexual, Transgender, Queer/Questioning, Intersex, Asexual/Aromantic/Agender plus additional subsects
MEP	Member of the European Parliament
MGTOW	Men Going Their Own Way
MRAs	Men's Rights Associations/Activists
NCIIS	Non-consensual intimate image-sharing
NCII	Non-consensual intimate images
PUAs	Pick-Up Artists
SRHR	Sexual and Reproductive Health and Rights
TFGBV	Technology-facilitated gender-based violence
TradCons	Traditional Christian Conservatives
Tradwives	Traditional wives
VLOPs	Very Large Online Platforms

Executive summary

Digital platforms have become central spaces for social interaction, participation and communication. They also play a central role in the spread of harmful and abusive content targeting women, girls and gender minorities. These issues form part of an increasingly interconnected ecosystem of actors, practices and ideologies undermining individual rights, reinforcing discriminatory gender norms and weakening democracy.

This working paper examines how online gendered harms and backlash against gender equality manifest, the role digital platforms play in shaping them, and the potential and limitations of the European Union's Digital Services Act (DSA) as a regulatory response. The analysis draws on academic and grey literature, as well as 17 semi-structured interviews with experts, policy-makers and civil society representatives across Europe. It argues that the DSA offers the EU's most significant framework to date for platform accountability and systemic risk governance, including in relation to gendered harms and gender backlash. Its effectiveness will depend on robust enforcement and the meaningful integration of gender responsiveness throughout its implementation.

Online gendered harms and gender backlash: four categories of analysis

The paper identifies four categories of online gendered harms and gender backlash. The first is gender-based hate and violence, which includes technology-facilitated gender-based violence (TFGBV), image-based sexual abuse and the activities of misogynistic online communities such as the 'Manosphere'. The second concerns the reinforcement of 'traditional' gender norms, including through 'Tradwife' and masculinity influencers who present restrictive, heteronormative gender roles as desirable or natural. The third is the suppression of gender equality content, voices and debate through opaque or biased platform moderation and the reduced visibility of feminist, LGBTQIA+ and sexual and reproductive health content. Coordinated attacks on gender equality advocates, activists and researchers constitute an important manifestation of this category. The fourth is gendered disinformation, which draws on false or misleading narratives and gender stereotypes to discredit women and gender minorities, particularly those engaged in public life. These categories are not exhaustive, nor are they mutually exclusive. A single harmful campaign may combine sexual harassment, gendered disinformation and the promotion of traditional gender roles.

The consequences of these online behaviours and forms of content extend beyond the digital environment. Targeted individuals or groups may experience psychological, economic, emotional and physical harm. Women politicians, journalists, researchers and activists are increasingly withdrawing from public debate due to online abuse and intimidation. Repeated exposure to misogynistic content can also normalise violence against women, girls and gender minorities,

as well as discriminatory attitudes in the physical world. The growing availability of generative AI on mainstream platforms adds further challenges related to scale, speed, responsibility and accountability.

How platforms shape online harms

Digital platforms are not neutral intermediaries. They are commercial actors whose advertising-driven business models reward user engagement, sustained attention and content virality. Algorithmic recommender systems privilege emotive, polarising and identity-based content, often exposing users to misogynistic material even when they have not actively sought it out, and reinforcing networks of like-minded users. Monetisation can create incentives for influencers who produce damaging, misleading or divisive content. These online dynamics help explain how narratives that were once considered fringe can become more visible, gain wider reach and gradually achieve broader social acceptance.

Other platform features also play a role. Modularity enables the rapid introduction of new services, including AI tools, sometimes without adequate safety and privacy safeguards. Anonymity can facilitate abuse and be exploited by perpetrators, creating additional challenges for accountability. Content moderation has been both insufficient in addressing gendered harms and coordinated harassment, and at times disproportionately restrictive towards gender equality-related content and voices. Uneven enforcement of platform policies across languages and contexts, limited opportunities for affected users to challenge moderation decisions and declining access to platform data for researchers and civil society organisations further inhibit accountability.

Opportunities for addressing gendered harms and gender backlash under the DSA

The DSA moves EU platform regulation beyond a predominantly liability-based approach towards broader obligations around due diligence, transparency and systemic risks. In force since February 2024, it introduces mechanisms to improve the identification and removal of illegal content and provide users with avenues to challenge arbitrary decisions. These include notice-and-action mechanisms, requirements for platforms to explain and review content moderation decisions, reporting obligations and provisions for Trusted Flaggers. The EU Directive on combating violence against women and domestic violence complements these provisions by establishing minimum criminal standards for several forms of online violence, including the non-consensual sharing of intimate images, cyberstalking and cyberharassment. The Directive must be transposed into Member States' national laws by June 2027.

The obligations the DSA places on Very Large Online Platforms (VLOPs) to assess and mitigate systemic risks are particularly relevant. Systemic risks explicitly include negative effects related to gender-based violence and extend to fundamental rights, civic discourse and electoral processes.

Unlike measures centred on individual instances of unlawful online behaviour, systemic risk obligations cover how a platform's architecture, including recommender systems, content moderation practices and design features, contributes to cumulative harms, including those that do not clearly meet the threshold of illegality. Independent audits, researchers' access to data and regulatory oversight can help expose these mechanisms and evaluate whether platforms' mitigation measures are adequate. Systemic risk obligations may also support greater scrutiny of gendered disinformation and the suppression of legitimate content related to gender equality.

Limitations of the DSA framework and challenges of enforcement

The DSA nevertheless has significant gaps. The legal definition of gendered harms remains ambiguous and varies across Member States, while many manifestations of gender backlash involve harmful but lawful forms of speech. Gendered disinformation is also not explicitly defined. The individual reporting systems introduced by the DSA can place a disproportionate burden on survivors, who may be required to repeatedly report content that reappears across multiple accounts and platforms. The DSA also does not adequately address the cross-platform nature of harmful content dissemination and the networks that facilitate it.

More fundamentally, the effectiveness of the DSA's systemic risk governance framework depends on the quality of platforms' self-assessments, the independence of external scrutiny and the capacity of public authorities to enforce obligations. To date, risk assessments have often been vague, particularly regarding gendered harms and gender backlash, raising concerns about the adequacy of both the evidence base and resulting mitigation measures. Civil society organisations and researchers have had limited involvement in risk assessments and consultations, while continuing to face significant constraints in terms of resources and access to data. These challenges are unfolding in a political environment marked by growing deregulatory pressures and a broader shift among major platforms away from commitments to safety, transparency and content moderation.

Policy priorities and recommendations

This analysis identifies five key priorities.

1. **First, EU and Member State authorities should strengthen DSA enforcement.** This includes systematically enforcing integration of gender-based violence, gendered disinformation and discriminatory content moderation and algorithmic systems into systemic risk assessments. It also requires the effective transposition of the EU Directive on combating violence against women and domestic violence, stronger coordination between EU and national authorities to enforce both regulations in conjunction and dedicated binding guidelines by the Commission.

2. **Second, EU and Member State authorities should fund and support civil society organisations and independent researchers, providing adequate resources, access to data and real opportunities to contribute to platform risk governance.** This includes participation in consultations and Trusted Flagger mechanisms, which should incorporate organisations and experts with specific gender expertise.
3. **Third, European policy-makers should resist pressure to weaken or prematurely simplify digital regulation** whose implementation is still developing and whose impacts have yet to be fully assessed.
4. **Fourth, regulators should leverage the DSA to place greater emphasis on the preventative and system safety measures platforms can take.** Platforms should be incentivised to adopt design and reporting features that give users more control over their online experience and reduce exposure to harmful content, while avoiding the transfer of safety responsibilities onto those encountering abuse.
5. **Fifth, given the transnational nature of digital platforms, the EU should integrate platform accountability and gender-responsive digital governance into its international partnerships, particularly the Global Gateway strategy, which has as one of its pillars investments in the digital sector.** The EU should ensure that investments and cooperation on digital technologies should not come at a price to gender and human rights, or amplify the risks of digital platforms. The EU should also engage in broader efforts to develop global standards for digital governance.

The DSA cannot, on its own, address the underlying social and political drivers of gendered harms and gender backlash. However, it can reshape some of the conditions through which digital platforms enable and amplify these dynamics, making it an important component of a broader response.

1 Introduction

Over the past decade, there has been a notable rise in harmful actions, content, communities and behaviours in online spaces, particularly targeting women, girls and gender minorities (European Institute for Gender Equality, 2022). This has contributed to psychological, mental health and even physical harm to individuals, and negative consequences at the level of society, including the silencing of marginalised voices, reduced participation of women and gender minorities in public life and the normalisation of extremist, misogynistic ideologies and hate speech.

The scale of the problem has grown considerably, in the number of people affected, the diversity of targets, the range of harmful behaviours and the speed at which harmful content and campaigns are amplified. Digital platforms bear a significant responsibility. Their business models, governance decisions and design features potentially allow, or even facilitate, gendered harms and backlash against gender equality.

Some jurisdictions have begun to adopt regulatory and legal measures in response. The European Union (EU) has been among the first to act in this area, adopting a comprehensive digital regulatory framework designed to mitigate some of the risks presented by online platforms and Artificial Intelligence (AI).

While the EU framework has been welcomed by advocates for rights, responsibility, ethics and online safety, it has faced a growing challenge from actors opposing digital regulation, including technology firms and their political allies. At the same time, a broader push for 'simplification' has permeated the digital policy debate in Brussels, despite the fact that implementation and enforcement of the EU's digital rules has only very recently begun.

This working paper contributes to current debates on digital regulation in the EU by examining how digital platforms shape different manifestations of gendered harms and backlash against gender equality, and whether and how regulatory responses can help address this. We focus in particular on the potential and limitations of the Digital Services Act (DSA). Considering that the DSA has been a model for others, we expect this study to help inform global debates as well.

The research is guided by three questions:

1. How are gendered harms and the gender backlash manifesting in online spaces (with a focus on Europe)?
2. What is the role of online platforms in relation to gendered harms and gender backlash?
3. What role can platform regulation, particularly the DSA, play in addressing these risks and harms?

The study draws on 17 semi-structured key informant interviews (KIIs) with academics, policy-makers and representatives from civil society organisations (CSOs) in Europe with expertise and experience in the topics covered by the research questions. The interviews were analysed in conjunction with relevant academic and grey literature. Throughout the report, key informant interviews are referenced using an interviewee number to ensure confidentiality.

The structure of the paper is as follows. Chapter 2 conceptualises digital platforms and Chapter 3 maps online gendered harms and backlash against gender equality. Chapter 4 examines the role of digital platforms in enabling online gendered harms and backlash against gender equality. Chapter 5 assesses the opportunities and limitations of the DSA in relation to gendered harms and gender backlash. The final chapter sets out recommendations for policy-makers.

2 Digital platforms

The term 'platform' refers to the applications and services offered by technology companies such as Facebook, Google Search and YouTube, as well as to the companies themselves (e.g. Meta, Microsoft) (Gorwa, 2019). Platforms are 'socio-technical infrastructures that mediate relationships among users' (Gillespie, 2018). They are not neutral intermediaries, but rather curate, rank and modulate visibility through algorithmic systems (see Box 1) designed primarily to maximise user engagement and economic return. Platforms and their algorithmic systems shape how online media acquires visibility, legitimacy and value.

Box 1 Algorithmic systems

Algorithms are encoded procedures for transforming input data into a desired output. Digital platforms rely on algorithmic systems based on computational procedures, data inputs and automated or semi-automated decision-making processes to rank, recommend, curate, moderate, personalise and target content and services. Algorithmic systems encompass algorithms, but also the datasets, machine-learning models, optimisation objectives and human oversight mechanisms that shape how decisions are generated and implemented.

Sources: Gillespie (2014); Beer (2017); Kitchin (2017); OECD (2022)

2.1 Three types of platforms

Three types of platforms are particularly relevant to this discussion: social media, closed messaging ecosystems and 'alt-tech' platforms.

Social media platforms are the most prominent or visible in shaping gendered harms and gender backlash online. They facilitate opportunistic interactions and selective self-presentation, 'either in real time or asynchronously, with both broad and narrow audiences who derive value from user-generated content and the perception of interaction with others' (Carr and Hayes, 2015). This definition excludes e-commerce platforms (like Amazon), or platforms that largely host content from a selective group of media producers (like Netflix). Most of the experts consulted for this research perceived large social media platforms like X, TikTok, Facebook, Instagram, Reddit and YouTube as the most prominent in shaping online gendered harms and gender backlash in Europe (KIIs 1, 2, 7, 11, 12). This is borne out by the literature, and evident from the sheer scale at which these platforms operate and the degree to which they mainstream content opposing gender equality and fostering gendered harms (Scheele, 2024).

Closed messaging ecosystems by design provide narrow channels of communication between individuals or within smaller groups, rather than allowing content that is broadly visible to the public at large, as mainstream social media platforms do. Telegram and Whatsapp are examples. As discussed in Chapter 4, platforms have the ability to integrate new features building on existing networks and architectures (this is known as modularity). This blurs the line between mass broadcasting and narrow systems, enabling the dissemination of harmful content across platforms.

'Alt-tech' platforms are smaller, ideologically permissive or purpose-built. They include platforms like Kick and Gab, specific, closed-channel Discord and Telegram servers, or other dedicated systems. These platforms are instrumental in radicalising users into more extreme spaces and ideologies (Collins et al., 2025; Gerrand et al., 2025).

2.2 Multiple dimensions of platforms

Platforms can be analysed from various dimensions. For the purpose of this paper, we highlight three: platforms as technologies; platforms as companies; and platforms as organisations. These three dimensions are inter-related and mutually constitutive. For example, the legal dimensions of platform regulation can influence how platforms govern online behaviour through their internal policies, content moderation and affordances.[1]

2.2.1 Platforms as technologies

Platforms comprise the software and hardware that provide distinct affordances to users – shaping the constraints and possibilities of online behaviour. Different user interfaces affect expectations and affordances for how the platform is used, including the media it hosts (text, audio or video), and how users interact between themselves, as well as with the central platform (including features to block or report content, or to determine what content one receives). Algorithmic systems curate media and determine the prioritisation and visibility of content within a network, as well as, in many cases, automatically flagging or censoring certain content, or targeting content to individuals based on assumptions about their preferences. Databases and identity systems model the behaviour of a network, profile users according to both pre-defined and constantly shifting data categories, including gendered categories, and shape how users find and interact with one another and with the broader network (for example, whether users can post anonymously).

---

1 Affordances refers to properties or features of platforms which affect the behaviour of users by constraining or enabling certain actions. It goes beyond the technical features of platforms to encompass what those features allow users to do, how they are used and how they shape online behaviours. Network association, social presence and anonymity are key affordances on social media platforms.

2.2.2 Platforms as companies

Platforms are commercial entities with distinct revenue models. Most commonly, platforms receive payments from advertisers for delivering targeted ads to users. In 2020, Facebook and Twitter generated $84.2 billion and $3.2 billion in advertising revenue respectively, representing 97.9% and 86.3% of their total revenue (Van der Vlist and Helmond, n.d.). Advertising makes up approximately 98% of Meta's revenue, 77% for TikTok and 68% for X (Science, Innovation and Technology Committee, 2025).

Advertising on most online social media works through the automated targeting of ad visibility towards highly specific data categories and niche user profiles, known as 'behavioural advertising' (Boerman et al., 2017). This means that the commercial infrastructure of platforms is, by construction, a machine for sorting users into pre-defined or inferred identity categories, including gender, and for delivering different content to different groups on that basis (Diepeveen, 2022). Advertiser targeting models are closely coupled with the algorithmic sorting and identity-based profiling features of platforms more broadly, and influences the content that is visible to different users.

The advertising-led revenue model favours 'clicks', 'views' and 'impressions'. This incentivises platforms to promote virality and engagement: to design content distribution and moderation systems that prioritise scale and visibility. This is often agnostic regarding the actual content of what is being promoted, or its potentially harmful nature. However, several studies have shown that content distribution and ranking that prioritises engagement systematically gravitates towards emotive and controversial content. This is particularly true for identity-based content (Milli et al., 2025; Rathje et al., 2021).

Platforms also support broader economic transactions and the so-called 'creator economy', primarily through revenue-sharing agreements and by allowing certain users to monetise content. Specific regulations govern what content is permitted, and the underlying economic infrastructure that supports (or constrains) content creators. Content policies can also include rules around demonetising particular kinds of content and disabling the commercial incentives for their production.

The economic considerations that influence organisational policy or governance mechanisms also encompass broader responses to shareholder priorities. These may include strategies to avoid regulatory scrutiny or pre-emptive bargains with political elites that go beyond short-term revenue models. Ownership structures influence corporate hierarchies, which can lead to certain actors having outsize influence over platforms' policies more broadly (Leerssen, 2025; Magalhães et al., 2026), including through ideologically driven policy-making.

2.2.3 Platforms as organisations

As organisations, platforms embody certain forms of distributed governance over a network, where a centralised 'core' establishes rules of behaviour in the network, while permitting different forms of interaction between heterogenous 'nodes' or members of the network.

The organisational structure of the platform determines how control is distributed between this 'core' and the broader network – from decentralised and 'bottom-up' governance systems, as exemplified by 'federated ecosystems' like BlueSky or Mastodon, to relatively centralised platforms like X, Meta or YouTube. The latter centralise decision-making around design features and policy decisions on content governance.

The 'core' unit of platforms is generally organised through the institutional form of a corporation. As such, platforms, as firms, are influenced by particular organisational dynamics of corporate governance which shape their influence on online behaviour, from media policies – such as guidelines on acceptable speech and user behaviour – to terms of service and the contractual entitlements and responsibilities between users and platform firms.

Internal policies

Platforms define and enforce internal policies applicable to users and communities interacting and engaging in the platform. Platforms' rules are meant to intermediate networked interactions and in principle balance different aspects and rights, such as free expression, safety, dignity and authenticity. For instance, so-called 'community standards' usually restrict content that could lead to harm, intimidation or violence – at least in theory.

Platforms also implement enforcement mechanisms linked with compliance with these policies, or other legal and regulatory requirements. These include differentiated enforcement responses to different kinds of content (e.g. banning, deprioritisation, amplification), or response mechanisms (e.g. human or automated moderation). As decentralised systems, platforms also inherently encompass a degree of devolution or decentralised governance.

Platforms' internal governance mechanisms determine their capacities and commitments towards accountability and transparency for harmful or illegal conduct. This includes mechanisms that allow users to report harmful content, to appeal moderation decisions and to understand why particular content gets removed, demoted or left in place.

As such, platform governance is characterised not only by the decisions made by the firm that controls the platform's core resources, but also overlapping commitments and structures, including decentralised content moderation mechanisms by user communities (such as community moderation on Reddit, or 'Trusted Flagger' programmes which work with expert groups to identify harmful content) (Singh, 2019).

Transparency

Independent scrutiny of platform systems by civil society actors, regulators and researchers depends on access to platform data. The level of transparency and access to relevant data provided by platforms is also linked to their governance decisions as organisations. Transparency has rapidly contracted over the last decade, making it more difficult for journalists, civil society organisations and researchers to monitor harmful content (Freelon et al., 2024). X's research access was restricted and repriced in 2023 and Meta shut down CrowdTangle in 2024.

The political economy of digital platforms

According to our consultations, over the last few years there have been rapid and significant changes in terms of compliance, accountability, transparency and collaboration by digital platforms related to preventing and addressing harmful online behaviour. This is apparently connected to the increasing alignment of major technology platforms' business interests and corporate policies with the ideological positions and deregulatory policies of the Trump administration. This manifests, for example, in reduced investments by platforms in trust and safety structures and teams, more permissive content moderation policies and less effective mechanisms to respond to users' claims.

One interviewee described how the ideological motivations of platform owners can influence online communities: '[*Platforms*] *are not neutral actors, but come from very specific visions of the words, very specific ideologies … we are seeing these clearly aligning with the extreme right*' (KII 15). Another noted how tech companies are not only prioritising 'profit over people, safety and equality in the design, deployment and use of their technologies', but are also providing largely unregulated platforms that enable networked activities which amplify and normalise far right views within mainstream discourse. These dynamics are frequently shielded by appeals to '*freedom of speech*', alongside lobbying aimed at dismantling existing regulatory frameworks and blocking new initiatives that could introduce limits on how platforms are used and how harmful activities emerge and spread (KII 17).

Since 2025, the US government and large US-based technology companies have also attempted to influence EU policy towards deregulation and flexible online content moderation. Measures have included diplomatic messaging, increased scrutiny by the US Congress, lobbying targeting the DSA, and pressure through trade tariffs and immigration restrictions (Lima-Strong and Jahangir, 2025).

3 Mapping gendered harms and backlash against gender equality

Online gendered harms and gender backlash comprise a range of online behaviours, content, campaigns and coordinated attacks promoting hate and violence against women, girls and gender minorities, or opposing gender equality and feminist gains. These can be illegal (e.g. hate speech or discrimination), or harmful but not illegal (e.g. discourse aggressively promoting 'traditional' gender norms or hostile to out-groups). While these online behaviours are increasingly coordinated by anti-gender actors and motivated by political ideologies, online violence against women can also be triggered by individual drivers – for example cyberbullying or the use of nudification applications to harass individuals for personal reasons. Examples of online gendered harms and gender backlash are described in Box 2.

Box 2 Case studies

Mass rape crimes against Gisèle Pélicot

In 2024, Dominique Pélicot was convicted by a French court for aggravated rape after more than a decade recruiting strangers to rape his then-wife Gisèle Pélicot following her sedation (Le Monde, 2026). Dominique Pélicot used a digital platform to recruit perpetrators and distribute non-consensual sexualised material. The same digital platform was also linked to cases of child sexual abuse and homophobic attacks (Gozzi, 2026).

Grok AI use to disseminate non-consensual sexualised deepfakes

Between late 2025 and early 2026, an Artificial Intelligence (AI) tool on X, known as Grok AI, created and disseminated large volumes of sexualised synthetic images, or non-consensual deepfakes, particularly targeting women and children. Within less than two weeks, Grok AI generated about three million of such sexualised images, including 23,000 appearing to depict children (DSA Civil Society Coordination Group, 2026). The EU and the UK both opened inquiries into the Grok AI sexual deepfakes (Booth and Gentleman, 2026).

Online campaign against gender equality in Europe

The ratification of the Istanbul convention on preventing and combating violence against women and domestic violence by several European countries (KII 2), and a 2013 report on Sexual and Reproductive Health and Rights (Estrela, 2013) drafted by the centre-left Portuguese Member of the European Parliament, Edite Estrela (KII 11), triggered an online backlash. The report was ultimately rejected in a plenary vote following a large, highly organised online campaign (Santos and Geva, 2022).

Online violence against women, girls and gender minorities often amplifies or is a precursor to violence in the physical world (European Institute for Gender Equality, 2022). Forms of violence facilitated by technology should be understood as part of a continuum between digital and physical aspects of harmful behaviour (Dunn, 2021).

The gender backlash is a form of political mobilisation that actively opposes gender equality and women's and LGBTQIA+ rights. It aims to roll back legal and social changes, often motivated by the belief that feminism has 'gone too far'. The online gender backlash represents an extension of this phenomenon to digital environments (see Box 3). In both online and physical environments, despite differences shaped by local contexts, what unites actors that oppose gender equality is the promotion of a version of society that is patriarchal, heteronormative and predominantly white, organised in a hierarchical and culturally, religiously and ethnically homogeneous way (Korolczuk et al., 2025). Gender-based violence in both online and physical environments has progressively become part of this phenomenon (Korolczuk et al., 2025).

Today's digital ecology is increasingly conducive to social harms (Gerrand et al., 2025), with research showing that misogynistic content has now become an expected part of social media. Nearly two-thirds of users encounter hate occasionally to regularly, with gender-based hate among the most prevalent forms (Reichelmann et al., 2021). In a poll of more than 3,000 Gen Z respondents from a nationally representative sample of the UK population, 73% have witnessed misogynistic content online, with half encountering it on a weekly basis (Amnesty International UK, 2025). Seventy per cent said that misogynistic and divisive language has increased on social media (Amnesty International UK, 2025).

Box 3 Online gender backlash

Online gender backlash refers to digitally mediated reactions against advances in gender equality. Harassment, abuse and silencing practices are used to punish perceived gender-norm 'transgressions' and reassert power hierarchies in online spaces. Such reactions are often coordinated and organised. They involve networks of transnational actors, shared narratives and the affordances of digital platforms. These factors enable collective and often large-scale targeting, amplification and sustained campaigns resisting or undermining gender equality and feminist progress.

Source: Tabbush (2025)

Our mapping identifies four categories and brings examples of each: gender-based hate and violence; reinforcement of 'traditional' gender norms; suppression of gender equality content, voices and debate; and gendered disinformation. While specific online manifestations help illustrate each of these categories, they are not mutually exclusive, meaning that one can contribute to others through feedback loops and spillover effects.[2]

3.1 Gender-based hate and violence

The most prominent category is the intentional, active, coordinated and harmful promotion of violence and hate towards women, gender minorities and feminism more broadly. Harmful behaviours and campaigns can target individuals, groups, organisations or whole categories, such as all women, feminists or LGBTQIA+ people. Here we look at two prominent examples: technology-facilitated gender-based violence and the Manosphere.

3.1.1 Technology-facilitated gender-based violence

Technology-facilitated gender-based violence (TFGBV) refers to acts of harm enabled or intensified through digital platforms, negatively affecting women's and LGBTQIA+ individuals' rights, safety and well-being (Tyabashe-Phume et al., 2025). TFGBV includes image-based sexual abuse (IBSA), stalking, bullying, sexual harassment, defamation, hate speech and exploitation. Tactics include non-consensual intimate image-sharing, deepfakes, sextortion, cyberstalking and doxxing.[3]

Box 4 Technology-facilitated gender-based violence

The International Center for Research on Women (ICRW) defines TFGBV as 'action by one or more people that harms others based on their sexual or gender identity or by enforcing harmful gender norms. This action is carried out using the internet and/or mobile technology and includes stalking, bullying, sexual harassment, defamation, hate speech and exploitation' (Hinson et al., 2018).

UN Women defines TFGBV as 'any act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or any other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms' (UN Women, 2025).

---

2 For example, the Manosphere simultaneously contributes to the first and second categories. We include it in the first category given that hate and violence appear to be a primary feature.

3 The term 'revenge porn' is now considered inaccurate and discriminatory. IBSA is the preferred term (Hörnle, 2026).

TFGBV affects millions of women, girls and gender minorities globally. In one survey of women aged 18–55 in Denmark, Italy, New Zealand, Poland, Spain, Sweden, the UK and the US (Amnesty International, 2017), an average of 23% – ranging from 16% in Italy to 33% in the US – reported at least one experience of online abuse or harassment. Research by UN Women in 12 countries in Eastern Europe and Central Asia found that more than 50% of women over 18 have experienced some form of TFGBV (UN Women Regional Office for Europe and Central Asia, 2023). Studies show a tendency to under-report such incidents (Inter-Parliamentary Union, 2018). Women who face intersectional forms of discrimination, including black and indigenous women, migrant women, women with disabilities and LGBTQIA+ people, face higher risks of TFGBV (Dunn, 2020; UN Women, 2023). Girls and younger women, who tend to use digital platforms more, are also more exposed.

Our consultation showed that all genders are affected by increasing violence and harm in digital environments, albeit to different extents. One expert emphasised that boys and young men are exposed to violent representations of what women supposedly represent to men:

> *Online pornography is getting more and more extreme and depicting violence against women, stuff which 20 years ago would have been regarded as obscene and illegal, and which would never be licensed in the sex cinema, but is now proliferating on the Internet.* [*This affects*] *in particular young boys … as young as 12, looking at this stuff and thinking this is normal behaviour against women, but it's actually violent and humiliating and goes far beyond just the sexual domain*. (KII 6)

3.1.2 The Manosphere

The Manosphere is a diverse ecosystem of online anti-feminist and male supremacist groups and influencers. They are united by the belief that mainstream society is systematically rigged against men and that gender policy is a 'war against men' (Gerrand et al., 2025). Discourse is characterised by misogynistic abuse, and incitement of violence against women is common.

Sub-groupings within the Manosphere include Men's Rights Associations/Activists (MRAs), Men Going Their Own Way (MGTOW), Involuntary Celibates (Incels), Pick-Up Artists (PUAs) and Traditional Christian Conservatives (Gerrand et al., 2025; Ignazzi et al., 2025).

MRAs promote the idea that men are victims of gender inequality, champion men's rights and consider that feminism has 'gone too far', to the point of marginalising men. The overarching narrative is that feminism has resulted in systemic discrimination against men. Commonly explored themes are child custody, divorce and workplace policies (Ignazzi et al., 2025).

Incels, or 'involuntary celibates', are men who perceive themselves as prevented from forming romantic or sexual relationships, which they blame on a societal preference for attractive or high-

status men and on women through dark humour, victimisation and the glorification of violence. Their perceived subordinate status is used to justify misogyny and the degradation of women (Ignazzi et al., 2025).

The MGTOW subgroup chooses to withdraw from relationships with women (Gerrand et al., 2025). Members promote a separatist, anti-feminist agenda focused on self-sufficiency and individual empowerment. Marriage and relationships are perceived as instruments used by women to subjugate men (Ignazzi et al., 2025).

PUAs focus on learning and applying techniques to supposedly attract and seduce women. They approach dating as a skill that can be optimised and controlled; sex is framed as conquest. PUAs are notably active on YouTube and other video-sharing platforms, providing 'guidance' on how to coerce women into sexual relationships (Ignazzi et al., 2025).

Traditional Christian Conservatives (often referred to as 'TradCons') reject casual sex and instead emphasise marriage and family structured around traditional lines. They promote a return to patriarchal, religiously justified roles for women and men. Feminism is framed as a social problem, to which traditional family structures are promoted as the solution. This community shares an anti-feminist ideology and stresses male authority, female subordination and the restoration of the heterosexual nuclear family (Ging, 2019). In many respects, the TradCons' ideology overlaps with the Tradwives, discussed below.

This online ecosystem of anti-women actors has experienced four key developments in recent years (Gerrand et al., 2025). First, the Manosphere has taken advantage of new social media platforms to both mainstream (via platforms like TikTok) and radicalise into more niche, underground spaces (such as Gab, Discord and Telegram). The format of short videos driven by recommender algorithms has been instrumental in pushing ideologically congenial and increasingly extreme content into users' feeds and rapidly amplifying content, as in the case of Andrew Tate's videos on TikTok. The incel community has grown in size and reach, and extremist discussion has increased (Baele et al., 2024). Our consultation pointed to the important role played by cross-platform migration of Manosphere-adjacent content. In particular, gaming spaces function as sites of misogynist indoctrination and there is significant crossover between male supremacism and online gaming communities (KIIs 6 and 9).

The second key development has been the Manosphere's monetisation and a shift towards wellness, which is linked to the rise of influencer culture. Key figures maintain and grow their audience to extract material, social or political resources. Their business models exploit the insecurities and vulnerabilities of their followers, while promising to deliver 'solutions' (Gerrand et al., 2025) through a focus on health and wellness.

The third development is the Manosphere's growing ideological and geographic diversity beyond the US (Gerrand et al., 2025). Although incel communities remain deeply racist and shaped by

white-supremacist hierarchies, Manosphere influencers have emerged in other regions. Examples include El Temach in Latin America (BBC Audio, 2026b) and Andrew Kibe in Kenya (BBC Audio, 2026a). Prominent Black figures, including Myron Gaines and Walter Weeks, also promote Manosphere content (Gerrand et al., 2025). The Manosphere has also become increasingly active in Europe (Ignazzi et al., 2025).

Finally, the Manosphere's boundaries have become increasingly porous, with cross-fertilisation among Manosphere groups, across platforms and with other extremist ideologies. Recent research shows how online misogyny functions as a gateway to far-right ideologies (Gerrand et al., 2025).

3.2 Reinforcement of 'traditional' gender norms

The second category reinforces 'traditional' gender norms and promotes a return to essentialised gender roles that frame men as breadwinners and women as home keepers, submissive to their husbands.

3.2.1 The Tradwives

The Tradwives (short for 'traditional wives') subculture exemplifies this trend. Tradwife influencers are (far)-right women promoting hyper-traditional gender roles on digital platforms like TikTok, Instagram and YouTube (Leidig, 2023). Five key features are commonly associated with Tradwives: anti-feminism, homemaking, adherence to traditional gender roles, affiliation with at least one position of extreme-right ideology (Conservative Right, Alt-Lite, Alt-Right), and an active presence as social media influencers (Sykes and Hopner, 2024).

Tradwives embrace biological differences between men and women, and advocate for gender roles based on a division between public and private life. According to this conception, men hold social and political power, and women are confined to the home as wives and mothers. Tradwives present themselves as full-time homemakers responsible for cooking, cleaning and childcare, while men act as primary breadwinners (Beaufils et al., 2025). This traditional arrangement is framed as a matter of 'personal choice'. Carefully curated online identities combine ideology and aesthetics to promote 'traditional' femininity while rejecting feminism (Sykes and Hopner, 2024). Content often evokes a nostalgic and idealised vision of 'traditional' femininity and motherhood, combined with a 'vintage off-grid cottage core lifestyle' (Beaufils et al., 2025).

Tradwives reject feminism and commonly describe themselves as 'feminine, but not feminist'. This framing is often used to advocate for a broader societal return to a conservative order, guided by religious and heteronormative beliefs about gender roles. Many portray their move away from feminism towards 'traditional' femininity as a personal 'awakening' (Sykes and Hopner, 2024). Feminist ideology is presented as undermining women's role within the family, and is frequently criticised as the reason why many women feel unable to be stay-at-home wives and mothers.

Tradwife influencers make considerable amounts of money out of their online activities (Sykes and Hopner, 2024). The 'stay at home' appeal also belies the reality of women's actual experience of economic dependence, heavy unpaid work burdens and reduced autonomy (Beaufils et al., 2025). As such, the Tradwife ideal functions not only as an aesthetic and cultural construct but also as a normative model that can obscure material inequalities and reinforce power imbalances driven by traditional gender norms, particularly through the promotion of financial dependence on a man and limited bargaining power within the household.

The majority of young women are not supportive of Tradwife content, and generally reject the idea of male decision-makers and female submission. What is potentially appealing in the Tradwife phenomenon is rather the relaxed leisure lifestyle its carefully curated videos present. This suggests that the appeal of Tradwives is more a symptom of the pressures facing young women juggling increasingly demanding jobs with intensive parenting responsibilities (Beaufils et al., 2025).

3.2.2 Influencers promoting stereotyped heteronormative masculinity ('manfluencers')

'Manfluencers' produce content promoting stereotypical heteronormative masculinity. Through topics such as fitness, financial success and relationships, they reinforce socially constructed models of masculinity framed as dominant and rooted in the assumption that heterosexuality and traditional gender roles are the natural, normal and superior way for men to behave.

Masculinity influencers have become mainstream. A 2025 report from Movember (Fisher et al., 2025) examining young men in the UK, the US and Australia found that nearly two thirds of surveyed individuals regularly engage with masculinity influencers. Many of those who follow this content describe self reliance and emotional control as important norms for men to embody. At the same time, the proliferation of narratives promoting heteronormative masculinity is also harmful to young men, who can face body-image pressures, sextortion, social isolation and psychological manipulation (Hayes et al., 2026).

Manfluencers may not necessarily attack women and feminism, but they illustrate the range of ways in which masculinity is tied to being straight, emotionally stoic, physically tough and explicitly non feminine. Young men engaging with this content tend to report more negative and restrictive attitudes towards women and their roles in relationships (Fisher et al., 2025). Rigid gender expectations that equate masculinity with dominance, sexual entitlement and emotional suppression can steer young men towards online misogyny (Hayes et al., 2026).

3.3 Suppression of gender equality content, voices and debate

Our consultation highlighted risks of opacity and discrimination against gender equality-related content, accounts and creators related to algorithmic moderation. While this is difficult to

demonstrate empirically, there is increasing evidence that content moderation decisions may not be applied equally for all types of users, and can lead to suppression related to people's gender, ethnicity or political orientation.

Emerging data suggests that many digital platforms, particularly through their governance frameworks and algorithmic decision-making, have effectively censored or reduced the visibility of marginalised communities and content that promotes feminist perspectives on issues including women's health (Uwazuruike, 2026), comprehensive sexuality education and abortion rights. One of our interviewees working with affected communities and topics told us:

> *I've been working as a sex educator and also creating erotic content online since 2016–17. There is a lot of censorship ...* [*it is*] *expanding to educators but also sex workers, queer content creators. Anything that is somehow non-conforming has experienced it, especially creators of colour, especially if you, in your appearance, don't conform to cis-heteronormative standards ... We call it censorship or discrimination.* (KII 1)

Another expert emphasised how platform policies and moderation systems can be restrictive in how they structure visibility and presentation, especially of LGBTQIA+ people and minorities online. A growing body of research shows that

> *platform users from queer communities, and who visibly present as queer, often are more likely to have that content removed or have it not be as visible or be suppressed somehow ... not only because they are from those groups, but also depending on how they're presenting themselves ... platforms are open to having more of what they would call 'family friendly', a sort of sanitised representation of LGBTQIA+ people* [*as being*] *'just like us'.* (KII 13)

One form of suppression is known as shadow banning. Kojah et al. (2025: 1) define shadow banning as 'a cross-platform light, insidious social media censorship technique whereby platforms avoid recommending or outright hide, content and profiles from their main discovery feeds, greatly affecting the visibility, earnings, mental health, and communications of the users affected'. Technically, platforms keep a user's content available, but through shadow banning their visibility is drastically reduced. This affects media, outlets, activists, entrepreneurs, and users who, in many cases, are not even aware that access to their content is being limited (Observacom, 2025). A report by the Observatorio Latinoamericano de Regulación, Medios y Convergencia (2025) highlights several examples where journalists, activists and communication or advocacy projects have reported a significant reduction in the reach, visibility or 'discoverability' of their content and accounts, without explanation or notification from the platforms concerned.

Haimson et al. (2021) find that content moderation, including shadow banning, has disproportionately affected marginalised communities, and in particular social media users who post nude content, who work as sex workers or who are part of the LGBTQIA+ community. The study found that, while conservative participants often had harmful content removed according to

platform guidelines on safe spaces and accurate information (e.g. regarding Covid-19), transgender participants' content was often removed as 'adult' despite following site guidelines, if critical of a dominant group (e.g. men, white people), or specifically related to transgender or queer issues. For black participants content removal was frequently related to racial justice or racism.

In addition to algorithmic restrictions affecting certain profiles, content covering topics deemed controversial (sex education, feminism, reproductive rights) has been subject to significantly lower distribution than other types of content, or even suppressed. The account 'Emma y yo' launched in 2019 as a response to the lack of accessible content on sex education in Latin America, reported a sharp drop in the reach of its posts, especially those related to sexual and reproductive rights. Through illustrations, infographics and educational materials, the project quickly became a regional reference for discussing sex and sexuality with children, adolescents and young adults. Gradually the account's visibility decreased, alongside changes to Meta's content policies. Other sex education accounts reported similar issues (Observacom, 2025).

A comprehensive study published in 2025 by the Center for Intimacy Justice (CIJ) found that sexual and reproductive health information for women and people of diverse genders has been systematically suppressed on digital platforms, while comparable content aimed at men has been permitted. Among the 159 non-profits, businesses and content creators surveyed, operating across more than 180 countries, a significant proportion reported facing substantial barriers when attempting to share information on topics such as menopause, menstrual health, endometriosis and other aspects of women's sexual and reproductive wellbeing (Center for Intimacy Justice, 2025). Following the release of the CIJ report, a coalition of women's health start-ups led by CensHERship filed formal complaints with the European Commission in March 2025. The complaints challenge major technology platforms over the systematic censorship of female health content. According to CensHERship, multiple cases have been documented, in which medically accurate, expert-led content has been blocked or labelled as 'adult' or 'political', while advertisements for male erectile dysfunction products continue to be approved (Tech.Eu, 2025).

Finally, our interviewees also mentioned that academics, researchers and scientists, in particular when engaging on gender equality debates, have been increasingly targeted by anti-gender campaigns and attacks online. This contributes to shrinking the online space for actual debate and accurate sharing of information.

Many interviewees mentioned that all women who have a voice and who are openly speaking about progressive topics can potentially become targets of campaigns motivated by hate, attempts of silencing and suppression, disinformation, among others (KIIs 3, 5, 11, 13, 15):

> *It's not just to women. The other thing about the gender backlash is when you talk about gender, any gender, any dimension of gender, there's a reaction. All genders receive it. I spend a*

> *lot of my time talking about women and violence against women online. But it's not just women. If you talk about the gender debate, even if you talk about gender without specifying a gender, there is a response and a reaction.* (KII 9).

Our consultation suggests that censorship, discrimination or invisibility have increasingly affected creators, content and debates deemed 'too feminist', 'too progressive', 'too liberal' or deviating from 'traditional' gender roles and norms. The affected circle seems to be expanding beyond groups usually most marginalised to a broader set of mainstream users and themes. As one interviewee put it, whereas platforms have been '*abusing their power to discriminate against creators*' and this has been '*tested on sex workers, queer creators, those who are at the margins*', there is a perception that '*censorship is expanding to people more in the mainstream*', and is being felt more significantly in the general public (KII 1).

3.4 Gendered disinformation

Disinformation refers to 'verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm' (European Commission, 2018). 'Gendered disinformation' can be understood as 'the spread of deceptive or inaccurate information and images against women political leaders, journalists and female public figures, following storylines that draw on misogyny and stereotypical gender roles' (Di Meco, 2020). Gehrke and Amit-Danhi (2025) define gendered disinformation as 'the employment of systematic and multidirectional flows of violence through (un)conscious content manipulation, audience engagement, and victim-blaming to prevent women and gender minorities from further political participation'. Predominant frames used in gendered disinformation campaigns include the narrative that women are unskilled, amateur and unfit for the job, and that women are subversive, manipulative, criminal and maliciously challenging of 'traditional' or 'natural' views of religion and social institutions, such as family and marriage (Gehrke, 2026). One common tactic is the recirculation of authentic photos out of their original context (Gehrke, 2026).

Gendered disinformation campaigns aim to silence, threaten or sideline female leaders from public engagement. One example often given is that of former Prime Minister of Finland Sanna Marin, who in 2022 was targeted by false and misleading online narratives, particularly following the release of videos of her dancing at a private party. Marin was a vocal supporter of Ukraine and effectively pushed for Finland to join NATO. Gendered attacks depicted her as 'irresponsible', 'unfit for office' or engaging in 'inappropriate behaviour', disproportionately emphasising elements not similarly central to attacks targeting male politicians, such as her young age, appearance, lifestyle and sexuality (Haukkamäki, 2025).

3.5 Implications

Online gendered harms and gender backlash have multidirectional and severely detrimental effects, reaching beyond the digital environment into the physical world. The most immediate consequences are experienced by victims themselves (Posetti and Shabbir, 2022), but they extend to audiences as well. The increasing circulation of harmful content, including on mainstream platforms, significantly contributes to the normalisation of violence, silencing, repression and subordination of women and gender minorities. The fact that such content is increasingly disseminated by 'celebrity influencers' and public figures (KIIs 3 and 16) helps legitimise attitudes, behaviours and positions that can be harmful (if not illegal) towards individuals and society at large. Among young men in particular, the circulation of narratives promoting the subjugation of women reinforces harmful perceptions of what women represent to men and what men are entitled to do to women and gender minorities. As one expert noted:

> *in countries like Italy and Spain … we passed from … high normalisation of violence, high normalisation of practices and especially of image-based sexual abuse towards a sort of revindication of this violence as a "right". So it seems that we're witnessing this anti-feminist backlash in which practices like image-based sexual abuse become a sort of a "right" among abusers.* (KII 15).

There is growing evidence that online violence has forced the retreat of women in the public eye, including politicians, from digital platforms or even from public life altogether (Michalko, 2026). This represents a significant threat to the quality and representativeness of democratic systems. Evidence also points to broader chilling effects, including self-censorship among those engaged in gender equality debates, including activists, journalists, academics, scientists and researchers (Posetti and Shabbir, 2022). Many no longer feel safe sharing accurate information meant to contribute to pluralistic perspectives.

Content creators at the margins, particularly LGBTQIA+ communities, feminist activists and those producing content on sexual and reproductive health and rights (SRHR) and women's health, seem to be disproportionately affected by opaque practices of content moderation and recommender systems, such as shadow banning, and coordinated hate campaigns. Affected users can experience adverse mental health outcomes and loss of income (KII 1). They can also feel under pressure to continuously adapt language and content in order to circumvent algorithmic bias. They can engage in self-censorship, or abandon social media altogether (Kojah et al., 2025).

The growing use of AI tools to create and disseminate synthetically generated harmful and gendered content is elevating the phenomenon of online gendered harms and gender backlash to a new level, as exemplified by the case of Grok (see Box 2). These AI tools enable harms to be generated, amplified and accelerated at even larger scale. They also raise complex questions regarding the veracity of circulated information, authorship and attribution of responsibility, as

well as accountability and punishment of perpetrators. These developments introduce additional regulatory challenges, underscoring the need for clearer responsibility and robust regulatory frameworks.

4 How platforms shape online gendered harms and gender backlash

Platforms do not determine social outcomes alone, but they increasingly shape online conduct and content through their technical and organisational systems. This happens in interaction with users' agency, system design, companies' business models and the broader legal and regulatory environment. While digital technologies and environments have not created gendered harms or gender backlash by themselves, they have significantly shaped, amplified and intensified manifestations of these phenomena. Scholarship on platforms' design features and data systems has demonstrated how categories like gender are 'baked in' to this infrastructure, since platforms either require users to declare a gender or infer one from behavioural signals, and reduce it to a small set of monetisable categories, often to drive advertisement-based targeting (Diepeveen, 2022; Bivens, 2017).

We add to this research by outlining here how four platform features have enabled gendered harms and online gender backlash. We also highlight how harmful content can travel across platforms and explore the interaction between technical and organisational aspects in relation to gendered harms and gender backlash.

4.1 Algorithmic systems and virality

Algorithmic systems are key to understanding how harmful and illegal content circulates and is amplified on online platforms. Algorithmic classification and targeting is based on personalisation, in which users are progressively shown more of what they, and users profiled as similar to them, have already engaged with. This happens through a common design principle called 'collaborative filtering' (Ekstrand et al., 2011). Recommendation systems are built on the principle of 'homophily', according to which similar users are assumed to want similar content. These systems connect users to others with ostensibly similar profiles and reduce exposure to different cultural, social or political perspectives (Haroon et al., 2023).

Algorithms prioritise engagement metrics over safety in their design. Coupled with identity mechanisms that connect users to others with ostensibly similar profiles, this leads to the amplification of extreme or borderline content, given that emotive, controversial and identity-based material generates stronger user response and engagement. Content expressing hostility towards an identity out-group is shared twice as often as other posts about the in-group (Rathje et al., 2021). Engagement-optimised systems are, in effect, calibrated to find and promote '*misogyny and any type of hatred*' (KII 7).

The role played by algorithms was further highlighted in our consultation: '*algorithms tend to amplify shocking, outrageous and polarising content* (...) *because* [*algorithmic governance*] *works*

on an engagement-based logic and wants to drive content that will get the most reactions out of people' (KII 7). This is also frequently leveraged by foreign interference and other actors to amplify and coordinate users with extremist political or ideological agendas, by using 'coordinated inauthentic behaviour' to simulate the engagement signals for media that ranking systems reward (Institute for Strategic Dialogue, 2020).

Through algorithms, platforms also allow ideologically aligned users to locate one another and consolidate into durable, self-reinforcing communities (Farvid et al., 2026). This has, for example, contributed to the spread of Manosphere and Tradwives content. Research on Tradwife influencers shows that platform algorithms are helping 'network resilience', by performing exactly as they were designed to: helping people within these networks find each other and keep their attention (Fitzgerald, 2020). One study assessing a series of fake TikTok and YouTube Shorts social media accounts to identify how algorithmic logics feed Manosphere content to boys and young men (Baker et al., 2024) found that, whether they searched for it or not, all of the male-identified accounts were recommended masculinist, anti-feminist and other extremist content, and the volume of such content increased rapidly once an account showed any interest by watching it. According to the study, removing the accounts of prominent misogynist influencers was largely ineffective, since their content remained among the most recommended through re-uploads by fans and smaller accounts, often routed through innocuous hashtags. Young people may not actively seek out this content, but rather are being algorithmically offered it. This is in turn being normalised by high levels of usage and generalised visibility (Baker et al., 2024).

Experimental research demonstrates the speed at which algorithmic amplification occurs. The Institute for Strategic Dialogue documented TikTok serving misogynistic and Manosphere content to accounts registered as belonging to teenage boys within short periods of ordinary use (Institute for Strategic Dialogue, n.d.). One UK study simulating typical usage by young teenagers found a fourfold increase in the level of misogynistic content presented on TikTok's 'For You' page after only five days of use (Regehr et al., 2024). The study found that the algorithm privileged progressively more extreme material, and that misogynistic ideologies were presented and gamified through soft or humorous cultural forms, blurring the boundary between entertainment and ideological content and lowering the threshold at which such material is accepted as ordinary. Algorithmic amplification may not always be intentional by platforms, but the superficially neutral design of algorithmic recommender systems still compromises online safety and amplifies gendered harms and gender backlash: platforms' apparently 'neutral' features are often, in practice, gendered (Schwartz and Neff, 2019). One expert interviewed noted that platform features promote virality which enables 'pile-ons' and 'band wagoning' (KII 9).

Our consultation also pointed to mechanisms used by content creators to maximise engagement and promote the virality of their content. Experts highlighted the prevalence of 'rage-baiting' [4] content and the propensity to virality within the Manosphere. As one policy analyst observed:

> *Much of what we would define as content that falls under the gender backlash category is inflammatory, it's sensationalist, it's controversial. So people will click, they will comment, they will tag. Platforms as amplifiers of all these kinds of things is very much inherent in their business model, because if it was all peaceful discourse, there would not be as much interest.* (KII 11)

A practitioner working with targets of online abuse described the same dynamic as a property of feed design:

> *The news feed as an affordance is designed to capture your attention. And virality, as part of the desire of the design of the news feed to capture attention, is putting the most trending information in the news feed or in the trending topics.* (KII 16)

Virality and algorithmic amplification are linked to coordinated harassment, particularly harassment of female politicians and other women with public profiles. Research on 'networked harassment' shows that attacks against women typically follow a recognisable sequence, in which an accusation amplified by a highly followed account licences a networked audience to attack, with platform metrics and recommendation systems supplying both the signal and the reward for participation (Marwick, 2021; Banet-Weiser and Miltner, 2016). One legal scholar interviewed described how individual reactions escalate into coordinated attacks through the interaction of bots and ranking systems:

> *As soon as you have any kind of comment on that, there will be a reaction from an audience or somebody who's seen it. And then, depending on the platform, some of the bots get involved. And once you've got the attention of some of those bots, there's almost an inevitable pile-on. And then it spreads and it becomes viral, and it only takes one thing to become viral. And then it's across platforms, and you're stuck in this loop of algorithmic repeat of that abuse.* (KII 9)

4.2 Modularity

Most platforms are 'modular' systems by design. This means that different technical components can be adapted and changed, often without disrupting the stability of the service. This architectural property allows platforms, for example, to incorporate features like AI content generation mechanisms into their services, as in the case of Grok or Telegram's Bot API, which allows users to automate and create their own content distribution system. Modularity implies

4 Rage bait is ordinarily considered a severe form of trolling that maliciously attempts to invoke other users into expressing outrage, thereby rewarding the user with engagement statistics that may drive income when monetised or some other reward such as the pleasure of attention

that a platform's design features can change rapidly and at scale, including where new features are incorporated across diverse operational contexts, or where safety features are removed. The speed and scale at which these changes are made make adaptation, including by users, regulators and researchers working on platform safety, challenging.

Interviewees noted how the integration of new features, such as GenAI systems, poses significant risks for various forms of gendered harms. Interviewees highlighted X's Grok image generation AI system, which led to a massive increase in image-based abuse, including the generation and sharing of non-consensual intimate images (NCII) or 'deepfakes', primarily targeting women. The use of AI has accelerated gendered harms at the same time as X has dismantled its internal content moderation safeguards, as well as deregulatory pressures around content governance enforcement (KII 2).

4.3 Anonymity

Anonymity creates a disinhibition effect (Suler, 2004) by enabling platform users to say or do things that are usually not acceptable in normal social interactions. It allows users to feel that their online actions are dissociated from actions in the physical world. Anonymity on online platforms like Telegram is a 'gendered affordance' (Semenzin and Bainotti, 2020), as it 'orients the diffusion of misogynistic practices and the performance of hegemonic masculinity' among online communities. Coupled with loose regulation, anonymity promotes a perception that the platform is a 'safe space' where individuals feel free to share, comment and denigrate female subjects without risks of social or legal consequences (Semenzin and Bainotti, 2020). That said, while some experts interviewed noted that anonymity on online platforms can contribute to cultures of impunity for online abuse (KII 16), others pointed out that it can enable marginalised groups to coordinate around shared concerns, and has been instrumental in enabling online press freedom (KII 4).

4.4 Moderation policies and enforcement

Platforms' internal content moderation practices are central to online safety. As one interviewee noted: '[Most] of the content that's taken down globally on platforms is because of a violation of their policies, not because it's illegal content' (KII 2). Platforms have, however, failed to invest appropriate resources towards content moderation against manifestations of gendered harms and gender backlash. As one interviewee explained:

> *There is certainly a lack of account moderation in the sense that we see lots of botnets publishing, distributing content that should totally be inappropriate in the light of any community rules that these platforms have … There is a legal requirement in Europe, for example through the Digital Services Act, through the Digital Markets Act for hard enforcement of community rules and moderation. But that's something that we currently do not see happening and … is certainly having huge impact on vulnerable communities.* (KII 5).

The Association for Progressive Communication notes that platforms' policies against disinformation rarely consider gendered disinformation as a distinct category of harm (Schnidrig, 2026).

The problem of inadequate moderation is particularly acute in relation to low-resource languages. Jang and Nanditha (2025) show how anti-feminist content in Hindi and Tamil circulated freely during the #MeToo backlash in India because moderation systems were calibrated to Western languages and harm categories. Memes and jokes mocking feminists were widely circulated on platforms such as X, Instagram and WhatsApp. These jokes, often in regional languages, typically trivialised the experiences of women involved in the #MeTooIndia movement, depicting them as liars, opportunists or part of a 'feminist conspiracy'. The use of regional languages meant that content often bypassed platform moderators, thriving in isolated, self-reinforcing communities that became echo chambers for anti-feminist rhetoric (Jang and Nanditha, 2025). In Europe, moderation resources are often inadequate for languages other than English. Interviewees expressed concern that 'certain languages aren't dealt with well on certain major platforms, like content moderation of languages, which aren't global majority languages is really poorly handled' (KII 9). Content moderation even for major languages like German and Portuguese was not sufficiently resourced (KII 5).

Conversely, as we have seen, platform policies tend to 'over-police' content and voices from gender minorities. Community standards have been criticised as reinforcing heteronormativity and exclusionary norms around nudity, gender and sexuality (Southerton et al., 2021; Griffin, 2025). Decisions on algorithmic prioritisation and amplification are the least accountable and transparent, since demotion and shadow banning are usually not disclosed or appealable.

Although community policies may exist in theory, there are gaps in how platforms enforce them. Gender minorities and other vulnerable groups have been particularly affected by the lack of response, action and content moderation:

> *The community policies around racism are not taken seriously. If you report an account, you get an automated response saying this wasn't against our terms and conditions. I think there's a real separation between theory and practice with platform legislation and platform governance. Policies that are unenforceable, codes of practise don't do anything. They are public relations, public affairs mechanisms.* (KII 16)

One recent study found that, from the perspective of survivors of gender-based violence, online platforms are usually slow and ineffective in responding to requests to remove abusive content. The study, focusing on victims in the US, found a general sense that content moderation is inadequate and that there is a disconnect between platforms' moderation policies and the reality faced by victims of violence, who repeatedly saw reposting of abusive material across multiple platforms (Equality Now and Sexual Violence Prevention Association, 2026). Reporting systems

typically permit complaints only against individual items of content, while gendered harassment arrives in coordinated peaks and across platforms, leaving targets to report an organised campaign one post at a time (Sinders, 2022).

4.5 Cross-platform dynamics

While online platforms work as separate entities with distinct technological and organisational features, in reality harmful content and user communities travel across platforms. Increasingly, actors perpetrating online gendered harms and gender backlash leverage features from different platforms to amplify harmful content across online spaces.

Closed messaging services such as Telegram can function as cross-platform infrastructure enabling gendered harm and gender backlash. Telegram combines strong privacy features – encrypted messaging, pseudonymous accounts and ephemeral content – with mass-distribution capabilities, supporting groups of up to 200,000 members and channels with unlimited audiences (AI Forensics, 2026). Telegram operates as a prominent redistribution hub where material scraped, leaked or captured from Instagram, Reddit or TikTok is aggregated, archived, repackaged and monetised (AI Forensics, 2026).

Research indicates that insular and extreme strands of communities central to gendered harms and the gender backlash tend to flourish on smaller forums and fringe platforms, where reduced moderation allows misogynistic and violent discourse to circulate more freely. Such spaces act as incubators in which narratives are developed and refined before migrating to mainstream or public-facing environments, where they influence wider online debates about gender and masculinity (Faith et al., 2026). Such cross-platform dissemination has played a key role in the virality and spread of gendered harms and gender backlash.

4.6 Interaction of technical and organisational factors

There is growing evidence that platforms' changing priorities have introduced more vulnerability related to gendered harms, linked to a more permissive political environment. At the same time, it is also becoming increasingly difficult to access the data required to build the evidence regarding gendered harms and gender backlash, audit whether recommender systems amplify misogynistic content, or evaluate whether platforms enforce their own policies. All this requires information and data that platforms increasingly withhold.

Since the acquisition of X by Elon Musk in 2022, a significant and sustained rise in hate speech has been observed on the platform (Hickey et al., 2025). The Center for Countering Digital Hate has documented the reinstatement of previously banned accounts, including prominent misogynist influencers like Andrew Tate (Center for Countering Digital Hate, 2022). Following the election of Donald Trump in 2025, large companies such as Meta rolled back safety commitments and

resources. Meta's reframed Hateful Conduct policy expressly permits allegations of 'mental illness' or 'abnormality' based on gendered or sexual orientation. This is an example of a clear governance decision with significant negative consequences for gender and sexual minorities.

Automated or algorithmic content moderation rely on human decision-making at various levels, whether in policy-level decisions about the classification of content, in the training of Machine Learning systems, or in human review for certain automated decisions. Choices regarding the nature of human involvement in content moderation matter. Platforms' content moderation and governance practices, including decisions on data access and transparency, often involve explicit policy decisions about what kind of content is acceptable, or how to prioritise and allocate resources for content moderation. Platforms are becoming more permissive, biased and opaque. As one interviewee noted:

> *There are people making decisions behind* [*algorithmic systems*] *and it could go from "it's too expensive to train algorithms to do a better job at discerning* [*what's good and bad content*]*", to "it's complicated to do a real taxonomy of what is erotic content, what is sex education and what is foreign* [*interference*]*" … it could range from that to intentional discrimination in algorithms and decisions.* (KII 9)

5 Platform regulation: potential and limitations of the Digital Services Act

With the DSA, the EU aims to better regulate online platforms through the application of due diligence, accountability, transparency and risk management obligations. Prior to the DSA, online platforms in Europe were governed by a patchwork of rules, particularly through the E-Commerce Directive (ECD). The ECD governs the responsibilities and liabilities of actors hosting, facilitating or publishing illegal information, while granting immunities to so-called 'intermediaries' that do not play an 'active role' in facilitating illegal acts online (Farrell, 2003; Klonick, 2017; Gillespie, 2018). The DSA significantly expands the scope and substance of platform regulation in the EU by harmonising the principles of intermediary liability established under the ECD. It also introduces new regulatory obligations for specific categories of digital services, including online platforms. In doing so, it substantially broadens platforms' legal responsibilities with respect to online safety and content governance, changing the regulatory approach from a liability-focused framework to a due diligence regime (Strowel and Meyere, 2023).

5.1 Main DSA requirements

The regulatory framework established by the DSA standardises how platforms should address illegal content and manage systemic risks associated with online behaviour and speech. While retaining and updating the ECD's approach to platform liability for illegal content, it introduces a wide range of additional obligations that platforms must comply with. Its goal is to ensure 'a safe, predictable and trusted online environment, addressing the dissemination of illegal content online and the societal risks that the dissemination of disinformation or other content may generate, where fundamental rights enshrined in the Charter [of Fundamental Rights of the European Union] are effectively protected' (Strowel and Meyere, 2023).

The DSA establishes tiered obligations according to a classification of platforms based on the nature and scope of services they provide. This ranges from minimal transparency and reporting obligations on 'intermediaries' and 'hosting providers' to more substantive obligations for 'platforms' and 'Very Large Online Platforms' (VLOPs). Online platforms are defined as services that principally offer 'a hosting service that, at the request of a recipient of the service, stores and disseminates information to the public' (Article 3). VLOPs are defined as online platforms with more than 45 million average monthly active users in the EU. These two categories include platforms most frequently related to the dissemination of gendered harms and gender backlash.

Two aspects of the DSA are particularly relevant to this discussion of gendered harms and gender backlash: general transparency and accountability obligations; and systemic risk obligations.

5.1.1 General transparency and accountability obligations

The DSA establishes mechanisms to make platforms generally more transparent and accountable, both to individual users and to the public. Article 16 requires hosting services, including platforms, to provide 'notice-and-action' mechanisms. These are online channels through which users can flag illegal content. Hosting services also have to inform the submitting user that the notice has been well received (Strowel and Meyere, 2023). A sufficiently precise and substantiated notice can give the provider actual knowledge of the illegality, and a provider that then fails to act expeditiously may be held liable as a publisher or facilitator of illegal action. Article 15 requires hosting services to issue information on those notices as part of their reporting obligations (Strowel and Meyere, 2023).

Beyond illegal content, Article 14 requires platforms to set out their content restrictions in their terms and conditions and to enforce them 'in a diligent, objective and proportionate manner'. Articles 17 and 20 set baselines for any restrictive decision, such as removal, demotion, demonetisation or suspension. Affected users must receive a statement of reasons and information on redress, and platforms must run an internal complaint-handling system open both to those who flagged content and to those subject to a decision.

At a public (rather than individual) level, Articles 15 and 24 require regular transparency reports on content moderation. These cover notices from users and Trusted Flaggers,[5] orders from national authorities and the accuracy and error rates of automated tools. Article 27 requires platforms to explain in plain language the main parameters of their algorithmic recommender systems and any options users have to influence them.

The DSA further establishes an ecosystem of independent actors with distinct responsibilities towards users and platforms. These include certified out-of-court dispute settlement bodies (Article 21), which resolve disputes over platforms' moderation decisions but cannot impose binding settlements. They also include 'Trusted Flaggers' (Article 22), designated expert organisations whose notices platforms must prioritise. Finally, users can mandate not-for-profit bodies to exercise their rights on their behalf, including through collective redress (Articles 86 and 90). Users can also lodge complaints with Digital Services Coordinators (Article 53) and claim compensation for damage caused by a provider's breach of its DSA obligations (Article 54).

---

5 Trusted flaggers are introduced by Article 22 as a status awarded by a Digital Services Coordinator to individuals with sufficient expertise and independence from online platforms. Notices sent out by these trusted flaggers within their area of expertise should be prioritised by online platforms (Strowel and Meyere, 2023).

5.1.2 Systemic risk obligations

The DSA subjects VLOPs to a distinct, risk-based regime, reflecting their outsize role in governing online behaviour and safety. Article 34 requires VLOPs to identify, assess and mitigate systemic risks arising from the design and functioning of their services, including their algorithmic recommender systems, content moderation and data governance practices. This must take into account both the probability and the severity of harms, including the rapid and widespread dissemination of illegal content or content that violates platform rules.

While the DSA does not define 'systemic' risk, the General Court in Amazon vs Commission (Case T-367/23) has characterised such risks as those arising 'to society as a whole', and that could affect a large part of the EU's population. The DSA's risk-based approach covers five main categories: (i) illegal content; (ii) negative effects on fundamental rights; (iii) negative effects on civic discourse and the electoral process; (iv) public security; and (v) negative effects in relation to gender-based violence, public health protection, the protection of minors, and serious negative consequences to a person's physical and mental well-being (van Hoboken et al., 2023).

Article 35 requires reasonable, proportionate and effective mitigation measures, tailored to the risks identified and mindful of their impact on fundamental rights. These may include adapting safety-focused design features, strengthening content moderation and cooperating with Trusted Flaggers and cross-platform coordination on online safety.

The risk management framework also includes transparency and oversight measures. Article 37 requires annual independent audits of VLOPs' compliance with the DSA's due diligence obligations, including risk assessment and mitigation. Article 40 gives regulators access to platform data to monitor compliance, and gives vetted researchers access to study systemic risks and the effectiveness of mitigation measures. Article 42 requires VLOPs to publish their risk assessment results, mitigation measures and audit reports once each audit is complete. Finally, Article 45 establishes a co-regulatory mechanism through which the European Commission can convene multi-stakeholder codes of conduct on illegal content and systemic risks. While participation is voluntary, the DSA requires an independent audit and assessment under Article 37, intended to provide indirect enforcement of codes linked to systemic risk obligations. The two principal codes currently operating under this provision are the Code of Conduct on Countering Illegal Hate Speech Online and the Code of Conduct on Disinformation.

5.2 Relevance to gendered harms and gender backlash

The DSA explicitly mentions platforms' responsibilities regarding illegal content, gender-based violence, the protection of fundamental rights and disinformation. While this does not directly address all forms of online gendered harms and gender backlash mapped in Chapter 3, its general

requirements on illegal content removal, platform transparency, risk assessment and mitigation provide important tools to tackle the four categories of gendered harms and gender backlash identified above.

5.2.1 Illegal content and coordination with the Directive on combating violence against women and domestic violence

The DSA requires platforms to remove illegal content when requested by relevant authorities. Illegal content is defined as 'any information, which, in itself or in relation to an activity, including the sale of products or the provision of services, is not in compliance with Union law or the law of any Member State, irrespective of the precise subject matter or nature of that law' (Strowel and Meyere, 2023). This means that certain harmful online behaviours and forms of content, namely gender-based violence and hate speech, may be addressed provided they are recognised as unlawful under EU or Member State legislation. However, the DSA stops short of providing standardised definitions for key concepts such as hate speech. As one interviewee observed: '*Illegal content is not really defined. And so hate speech is also not really defined under illegal content … which makes it difficult … to hold perpetrators or the platform accountable*' (KII 3).

Following the adoption of the DSA, the EU approved the Directive on combating violence against women and domestic violence, published in May 2024. The Directive establishes minimum criminal standards for several forms of digital violence, including the non-consensual sharing of intimate images, cyberstalking, cyberharassment and cyber incitement to hatred (KIIs 2 and 10). As one expert noted, '*the DSA is an enormous tool, but in particular on gender-based violence, the way the directive* [*on violence against women and domestic violence*] *and the DSA are going to work together could fundamentally change our tech-facilitated gender-based violence landscape*' (KII 2).

Together, the two instruments provide an important mechanism for addressing harms arising from illegal gendered content and online behaviour, particularly from the perspective of fundamental rights protection and criminal law enforcement. While the DSA requires platforms to establish mechanisms for identifying and removing illegal content, the Directive clarifies which forms of gender-based violence and abuse are unlawful across the EU. By harmonising legal definitions and criminalising key forms of gender-based violence across Member States, including manifestations in the digital sphere, the Directive enhances implementation of the DSA in this area.

Experts have also noted the role of Trusted Flaggers under the DSA in enabling the expeditious removal of illegal content linked to gender-based violence owing to the priority given to processing their notices (Centre for Democracy and Technology, 2024). While gender dimensions are not highlighted in the current designation of Trusted Flaggers, the EU Commission has indicated in its Gender Equality Strategy (2026–2030) that it will support build capacity of Trusted Flaggers on gender-based cyberviolence, by clarifying their role in Guidelines on Trusted Flaggers to be issued.

5.2.2 Transparency and due diligence obligations

Transparency and due diligence requirements in the DSA can potentially help address manifestations of gendered harms and gender backlash. Interviewees noted in particular that Article 40 enables researchers to better understand and shape systemic risk assessments, a crucial tool to ensure independent accountability of platform practices by academics and civil society organisations. Greater transparency around algorithmic recommender systems, for example, can help researchers, civil society and regulators scrutinise how platforms amplify, distribute and make visible Tradwife, Manosphere and 'Manfluencer' content. Mechanisms to make recommender systems more transparent, alongside provisions allowing affected parties insight into, and the ability to challenge, content moderation decisions, can help ensure that opaque practices do not arbitrarily restrict freedom of expression in ways that reinforce harmful gender norms and contribute to the gender backlash (Appelman, 2023). Interviewees noted, for example, that the DSA's appellate processes against content moderation decisions could provide a way to challenge demonetisation or unfair removals (KII 1).

5.2.3 Gender-based violence and gendered disinformation as systemic risks

The DSA's systemic risk approach creates a specific framework to address more structural, diffuse harms that are not clearly linked to individual content instances or criminal law. It allows for intervention in the design and governance of platforms in ways that can account for the more structural nature of these harms (Suzor et al., 2019). The DSA's provisions on systemic risk governance by VLOPs require that providers of VLOPs and VLOSEs must assess 'any actual or foreseeable 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 wellbeing'. As highlighted by one interviewee:

> *the inclusion of gender-based violence in the Digital Services Act provides us with the most substantive opportunity that we have to create truly intersectional frameworks for how we deal with platform governance, because there's so many elements of it that are just not going to fall under illegal content, but that's having a clear impact on democratic participation.* (KII 2)

The gendered dimensions of harmful online content as systemic to online platforms, rather than just incidental, is increasingly recognised by institutions responsible for steering and enforcing the DSA. The European Board for Digital Services, established as a coordination mechanism under the DSA, noted in its first report on systemic risks that online services covered under the DSA have 'contributed to the emergence and amplification of gender-based cyber violence' as well as linked risks of illegal content and discrimination. It also notes that 'while systemic risks related to GBV were identified across all types of services, they were most frequently reported by providers of social media platforms. Providers and CSOs highlighted that GBV disproportionately affects women and girls, especially those from marginalised groups, including ethnic or religious minorities, people with disabilities, and LGBTQIA+ communities' (European Board for Digital

Services, 2025). In early 2026, the Commission opened an investigation into whether X breached its obligations to mitigate systemic risks posed by its recommender systems and its integration with Grok AI (European Commission, 2026).

The DSA may also provide a framework for addressing gendered disinformation, particularly through transparency and risk-management obligations related to disinformation more broadly, and risks concerning its negative effects on democratic processes, civic discourse and electoral processes, as well as public security. One of the objectives of the DSA is to address the spread of disinformation, which appears in several recitals. However, the regulation lacks a definition of the term (Strowel and Meyere, 2023). Transparency obligations on platforms could help curb the dissemination of disinformation. For instance, the DSA requests VLOPs to 'focus on the information which is not illegal, but contributes to the systemic risks identified in this Regulation. Providers should therefore pay particular attention on how their services are used to disseminate or amplify misleading or deceptive content, including disinformation' (Strowel and Meyere, 2023). Risk mitigation measures for some situations that directly relate to disinformation, such as the circulation of deep fakes, have also been included in the Regulation (Strowel and Meyere, 2023).

5.3 Limitations of the DSA in addressing gendered harms and gender backlash

The DSA provides important safeguards and measures to potentially mitigate gendered harms and gender backlash, but there remain important areas of ambiguity, gaps and limitations, notably around enforcement.

5.3.1 Ambiguities and enforcement challenges

Our consultations noted that ambiguities around the legal nature of different forms of online gender-based violence and manifestations of gendered harms and gender backlash can make notice-and-action tools less reliable, particularly as the question of legality around many of these issues is determined by individual Member States. This leads to unclear thresholds for determinations of illegality. Similarly, notice-and-action requirements to enforce platforms' own policies, terms and conditions, or community guidelines on harmful speech, are often vague, and the scope of their actual enforcement under the regulation remains unclear (Leerssen et al., 2025).

Beyond ambiguity issues, measures such as notice-and-action shift the burden of monitoring, reporting and ensuring that platforms take action against illegal speech onto survivors of harm. These actors end up being held responsible for managing the risks of abuse, including finding and reporting it (McGinley and Woods, n.d.). This is a particular concern given that content, accounts or networks can reappear after being removed, either on the same platform or on others (Ding et al., 2026).

Interviewees also highlighted how the DSA's notice-and-action mechanisms by design were not '*user friendly or trauma informed*' (KII 2), with one interviewee describing their experience working with survivors of online abuse as follows:

> *none of the survivors we interviewed could say that they got any sort of satisfactory resolution through the use of the platforms, including even through the use of the law … the reporting is opaque. It's not very clear.* (KII 17).

Another said:

> *the lack of proactive enforcement means that they will only act on user reports, and we know how complex and tedious it can sometimes be to report content on platforms that don't necessarily make it easy. You don't always get a response.* (KII 7)

Reporting mechanisms have been criticised as forcing users to attribute their experiences to predetermined categories which fail to capture the multifaceted nature of the abuse (Allen, 2022).

While the Directive has been a crucial step in harmonising gender-based violence, its transposition by Member States will be an entirely new challenge in itself, together with the considerable coordination needed between Member States and EU authorities to enforce both the DSA and the Directive in conjunction.

5.3.2 Insufficient integration of gendered harms in the scope of systemic risks

Online gendered harms and gender backlash manifest in ways that are not necessarily or only restricted to illegal speech or criminalised behaviour. As such, the DSA's more holistic approach represents a significant departure from previous regulatory mechanisms, which focused largely on removing illegal speech. At the same time, whether the systemic risk framework is sufficient to address such phenomena is unclear, with several interviewees pointing to the continued importance of criminal justice frameworks in addressing online gender-based violence (KII 2, KII 3, KII 6, KII 16).

The European Centre for Not-for-Profit Law (ECNL) notes five significant limitations of systemic risk reporting from major platforms: vague risk statements rather than detailed, evidenced assessments; insufficient attention to fundamental rights relative to other systemic risk categories; irrelevant or missing data, with claims about risks or mitigations often unsupported by evidence; limited stakeholder engagement, leaving it unclear how civil society or affected-community consultation actually shaped risk identification or mitigation; and assessments skewed towards English-language and Western European contexts (ECNL, 2026).

In March 2025, the DSA Civil Society Coordination Group, an independent group convened by the Centre for Democracy and Technology, evaluated the first set of risk assessment reports

published by VLOPs, documenting their practices and mitigation measures. This found that, while platforms like X and Meta documented risk management measures through both their design (such as blocking or muting functions) and policy changes, they were at the same time removing measures preventing online gender-based violence (DSA Civil Society Coordination Group, 2025). Apart from the specific mitigations put in place for online gender-based violence, systemic risks arising from gender backlash are also risks to fundamental rights and non-discrimination, as well as civic discourse and public security. These inter-connected systemic risks, including risks from gendered disinformation, or systemic risks arising from the use of AI 'nudifying apps', have not been considered by VLOPs in their systemic risk assessments (Women's Aid, 2026).

Another limitation in the DSA's design is its relative lack of regulation of the cross-platform spread of illegal and harmful content. As noted in the previous chapter, harmful content travels across platforms, from creation to dissemination to interaction, often through organised networks. Cross-platform harms are not acknowledged in the DSA directly, and a study on platforms' responses to coordinated foreign interference indicates that they do not necessarily account for cross-platform behaviour in their assessments of systemic risk (DG Connect, 2023).

5.3.3 Inadequate resourcing and consultation with civil society actors

The DSA specifically notes that VLOPs must 'conduct their risk assessments and design their risk mitigation measures with the involvement of representatives of the recipients of the service, representatives of groups potentially impacted by their services, independent experts and civil society organisations'. However, interviewees noted that civil society participation in DSA systemic risk governance is structurally uneven and under-resourced, with limited coordination between civil society groups and researchers (Carvalho and Griffin, 2026). One expert noted that '*CSOs don't get the impression that they are backed by the European Commission*' (KII 7).

Interviewees also noted that the DSA's mechanisms for accountability were premised on the ability of institutions both internal and external to the EU (including civil society) to evaluate and take action (KII 1, KII 2). These institutions are critically under-resourced. One interviewee remarked: '*The European Institute for Gender Equality has been designated one and a half persons to help them with the implementation of the* [*Directive on Violence Against Women and Domestic Violence*] *across the entire European Union*' (KII 2). Organisations with expertise on gender equality may not have the capacity or resources to become certified as Trusted Flaggers, a reality acknowledged by the Commission in its guidelines (Center for Democracy and Technology, 2024).

5.3.4 Co-regulation and compliance challenges

The systemic risks identification and mitigation mechanisms established by the DSA rely on the cooperation of platforms, in what is described as a 'co-regulatory regime' (Husovec, 2024). This potentially allows regulated entities to design risk assessments and mitigation mechanisms in ways that are minimally compliant and that favour their own interests (Griffin, 2025; Palumbo, 2026).

Moreover, as Rigotti et al. (2026) argue, the DSA risk assessment framework, 'while analytically useful, struggles to capture harms that are cumulative, structurally gendered, or not reducible to discrete quantifiable incidents'.

In the absence of adequate external checks and balances, there is a risk that platforms treat risk assessment and management as a check-box exercise, where they both define the scope of risks they account for, and design the minimal viable interventions to counter them. Regarding the standardisation of regulatory exercises and risk management, one interviewee noted:

> *we've shifted from companies feeling incentivised to show that they respect human rights due diligence to this is now just a compliance exercise. So we're going to have almost a race to the bottom of what we legally need to do to comply with the law … The DSA kind of just codified what platforms were already doing in the first place, and when it was concluded, platforms were actually celebrating the DSA for taking a co-regulatory approach.* (KII 2)

Consulted experts from civil society noted that platforms may also not be following the spirit of the DSA's consultative mandates:

> *platforms should consult with civil society organisations and impacted groups, and thus far, amongst our big networks that we know, that's not happening. What happens is that they turn up to multi-stakeholder events and then say that was the moment in which they did consultation … or they present us their risk assessments after the risk assessment has been done. And at that point we can't inform* [*the*] *conclusions.* (KII 2)

Another interviewee emphasised that existing platform engagements with civil society or affected people do not acknowledge the power imbalance between the platform company (as a well-resourced organisation with information about the platform) and users or civil society representatives. In the absence of structured dialogue, platforms set the terms of engagement with civil society and affected users (KII 4).

6 Policy priorities and recommendations

Online gendered harms and gender backlash comprise a range of behaviours, content, campaigns and coordinated attacks promoting hate and violence against women, girls and gender minorities, and opposing gender equality gains. These can be illegal (e.g. hate speech or discrimination), or harmful but not illegal (e.g. discourse aggressively promoting 'traditional' gender norms or hostile to out-groups). Four categories are mapped in this paper: gender-based hate and violence; reinforcement of 'traditional' gender norms; suppression of gender equality content, voices and debate; and gendered disinformation. The scale of these challenges has grown considerably in recent years, in the number of victims affected, the diversity of targets, the range of harmful behaviours and the speed at which harmful content and campaigns are amplified. The integration of AI tools in mainstream digital platforms aggravates the issue.

Digital platforms bear a significant responsibility, and research demonstrates that their business models, governance decisions and design features potentially allow, or even facilitate, gendered harms and backlash against gender equality. Governments have developed policies and regulations to hold technology companies to account in response to the negative impacts that digital platforms increasingly have on societies and individuals. The DSA is the most comprehensive framework to date to address online gendered harms and gender backlash. This is notably the case regarding requirements for platforms to remove illegal content, alongside broader transparency, accountability and risk mitigation requirements.

Much depends on enforcement by DSA authorities, compliance by platforms, and the strength of the co-regulatory system, including civil society and researchers. Enforcement of the DSA has been limited by insufficient resourcing of enforcement authorities and other relevant co-regulatory actors, in particular the lack of gender expertise within enforcement authorities and ineffective coordination with gender equality organisations. Increasing deregulatory pressure, including from powerful and well-resourced platform firms but also from within the EU, could contribute to ineffective enforcement as well as looser regulatory standards.

Considering the DSA's role in addressing online gendered harms and gender backlash, this analysis identifies five key priorities.

6.1 Strengthen DSA enforcement

First, EU and Member State authorities should strengthen DSA enforcement. This includes systematically integrating gender-based violence, gendered disinformation and discriminatory

content moderation and algorithmic systems into systemic risk assessments. These areas are already within the scope of the DSA, but have not been adequately integrated in platforms' risk assessments.

The effective transposition of the EU Directive on combating violence against women and domestic violence will also be essential, particularly given the risk of uneven interpretation of gender-related offences and disparities in enforcement capacity across the EU. Strong coordination between EU and national authorities, both in transposing the Directive and in enforcing the two regulatory frameworks in a coherent manner, will be crucial. The European Commission should develop dedicated binding guidance, in close consultation with digital rights and gender equality organisations, as well as researchers, to support consistency in this process.

6.2 Resource civil society to support enforcement

Second, civil society organisations and independent researchers need adequate resources, access to data and meaningful opportunities to contribute to platform risk governance. This includes participation in consultations and Trusted Flagger mechanisms, which should incorporate organisations and experts with specific gender expertise.

Our consultations highlighted that, despite being essential actors in the co-regulation of platform governance, civil society organisations and researchers have increasingly been sidelined both by EU institutions and by platforms. Yet their expertise is critical to monitoring risks, building the evidence base and holding platforms accountable. There remains a significant imbalance in resources and capacity between platforms on the one hand, and regulators, researchers and civil society on the other. To help address this, enforcement authorities, digital rights organisations and researchers should work more closely together, pooling resources and expertise to strengthen oversight and improve accountability.

6.3 Resist de-regulatory pressure

Third, European policy-makers should resist pressure to weaken or prematurely simplify digital regulation whose implementation is still developing and whose impacts have yet to be fully assessed.

Our consultations noted that US-based tech companies have exerted increasing pressure within the EU, often backed by the Trump administration, to weaken and 'simplify' digital regulation, under the guise of 'defence of freedom of speech' and constraints to competitiveness and innovation. The EU must resist pressure to weaken online safety protections under the DSA and prioritise citizens' safety, privacy and human rights, as well as the universal, egalitarian values that have underpinned the European project. Framing innovation and free expression in opposition to

rights and online safety ignores the fact that the burden of the absence of regulation is not only borne by women, gender and sexual minorities, but by society as a whole, including by amplifying security risks related to radicalisation and extremism.

The DSA's obligations relating to non-discrimination and fundamental rights should not be dismissed as administrative burdens to be streamlined alongside procedural requirements, particularly given that the provisions relevant to countering the online gender backlash have yet to be systematically analysed or studied. Reducing these obligations before they have been implemented would foreclose any assessment of their effectiveness and their benefits in reducing systemic societal risks.

6.4 Emphasise preventative measures and empower users

Fourth, regulators should leverage the DSA to place greater emphasis on the preventative measures platforms can take. Regulators should encourage platforms to adopt design and reporting features that give users greater control over the content they see and their online experiences, helping to reduce exposure to harmful content while avoiding transferring responsibility for safety onto those experiencing abuse.

Given platforms' limited compliance and cooperation in several DSA-mandated areas, particularly regarding transparency around algorithmic systems and content moderation policies, DSA authorities should broaden their approach and place greater emphasis on preventative measures. Recent policy developments in Australia will require platforms to provide users with clearer choices and greater control over their online experiences. New legislation under development aims to place greater responsibility on digital service providers to ensure a safer online environment by establishing minimum standards for the features and tools they deploy (Albanese and Wells, 2026). The proposed legislation includes the initiative 'My Feed, My Way', which seeks to give users greater control over their social media algorithms. Such initiatives could serve as inspiration for EU policymakers.

6.5 Integrate platform regulation aspects in international partnerships

Fifth, the EU should integrate platform accountability and gender-responsive digital governance into its international partnerships, particularly Global Gateway. The EU should also engage in broader efforts to develop global standards for digital governance.

There is scope for further and more effective international partnerships promoting collaboration, global standards and best practices in digital policy and regulatory enforcement, particularly given the transnational nature of digital platforms' businesses and use, and the limited reach of territorial-specific regulations.

While much has been said about the so-called 'Brussels effect' in terms of exporting of EU's digital regulation framework, there is still significant scope for the EU to integrate digital regulation standards and platform accountability practices in its international partnerships. Mechanisms promoting climate, human rights, democracy and gender equality standards tend to be integrated into the EU's development cooperation. The EU should also consider adopting digital regulation and enforcement counterparts or conditionality measures in its partnerships. This is particularly relevant in the context of the Global Gateway strategy, which aims to support partner countries to develop connectivity and digital infrastructure. These investments should not come at the price of further amplifying the potential damage and risks posed by digital platforms.

The DSA cannot, on its own, address the underlying social and political drivers of gendered harms and gender backlash. However, it can reshape some of the key conditions through which digital platforms enable and amplify these dynamics, making it an important component of a broader response.

References

AI Forensics (2026) *Harassment as Infrastructure: How Telegram's Design Enables TFGBV*. https://aiforensics.org/work/telegram-harassment-infrastructure.

Albanese, A. and Wells, A. (2026) My Feed, My Way. Press release (www.pm.gov.au/media/my-feed-my-way).

Alcantara, J. and Valentim, J. (2023) 'Gender-Based Disinformation: A Scoping Review of the Literature, 2013–2023'. Ex Æquo - Revista Da Associação Portuguesa De Estudos Sobre As Mulheres, Dossier 'Discurso de ódio misógino: representações, impactos e intervenções', no. 48 (https://exaequo-ojs.apem-estudos.org/exaequo/en/article/view/114).

Allen, A. (2022) 'An Intersectional Lens on Online Gender Based Violence and the Digital Services Act'. Verfassungsblog (https://doi.org/10.17176/20221101-215626-0).

Amnesty International (2017) 'Amnesty Reveals Alarming Impact of Online Abuse against Women'. Press release (www.amnesty.org/en/latest/press-release/2017/11/amnesty-reveals-alarming-impact-of-online-abuse-against-women).

Amnesty International UK (2025) *Gen Z in the UK: Perceptions and experiences of misogyny on social media*. Savanta report (https://media.amnesty.org.uk/documents/Amnesty20International20-20Gen20Z20and20Online20Misogyny20202420-20Savant_15tE7g9.pdf).

Appelman, N. (2023) *Disparate Content Moderation*. Institute for Information Law, University of Amsterdam (http://dsa-observatory.eu/2023/10/31/research-report-on-disparate-content-moderation).

Are, C. (2025) 'Flagging as a Silencing Tool: Exploring the Relationship between de-Platforming of Sex and Online Abuse on Instagram and TikTok'. *New Media & Society* 27(6): 3577–95.

Baele, S., Brace, L. and Ging, D. (2024) 'A Diachronic Cross-Platforms Analysis of Violent Extremist Language in the Incel Online Ecosystem'. *Terrorism and Political Violence* 36(3): 382–405 (https://doi.org/https://doi.org/10.1080/09546553.2022.2161373).

Baker, C., Ging, D. and Brandt Andreasen, M. (2024) *Recommending Toxicity: The Role of Algorithmic Recommender Functions on YouTube Shorts and TikTok in Promoting Male Supremacist Influencers*. Dublin City University Anti-Bullying Centre. (www.dcu.ie/antibullyingcentre/recommending-toxicity).

Banet-Weiser, S. and Miltner, K.M. (2016) '# MasculinitySoFragile: Culture, Structure, and Networked Misogyny'. *Feminist Media Studies* 16(1): 171–74.

BBC Audio (2026a) 'Manosphere Messiahs: Kenya'. Radio programme, 8 June (www.bbc.co.uk/programmes/w3ct9b95).

BBC Audio (2026b) 'Manosphere Messiahs: Mexico'. Radio programme, 28 May (www.bbc.com/audio/play/ponnly27).

Beaufils, C., Chung, H. and Yuan, S. (2025) *Tradwife: Between Myths and Realities*. King's Global Institute for Women's Leadership Report. King's College London.

Beer, D. (2017) 'The Social Power of Algorithms'. *Information, Communication & Society* 20(1): 1–13 (https://doi.org/10.1080/1369118X.2016.1216147).

Bivens, R. (2017) 'The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook'. *New Media & Society* 19(6): 880–98 (https://doi.org/10.1177/1461444815621527).

Boerman, S.C., Kruikemeier, S. and Zuiderveen Borgesius, F.J. (2017) 'Online Behavioral Advertising: A Literature Review and Research Agenda'. *Journal of Advertising* 46(3): 363–76 (https://doi.org/10.1080/00913367.2017.1339368).

Booth, R. and Gentleman, A. (2026) 'Wave of Grok AI Fake Images of Women and Girls Appalling, Says UK Minister'. The Guardian, 6 January (www.theguardian.com/technology/2026/jan/06/grok-ai-fake-images-women-girls-undressed-uk-minister-liz-kendall).

Busch, C. (2022) 'Regulating the Expanding Content Moderation Universe: A European Perspective on Infrastructure Moderation'. *UCLA JL & Tech*. 27: 32.

Carr, C.T. and Hayes, R.A. (2015) 'Social Media: Defining, Developing, and Divining'. *Atlantic Journal of Communication* 23(1): 46–65 (https://doi.org/10.1080/15456870.2015.972282).

de Carvalho, M.C. and Griffin, R. (2026) *Who Speaks and Who Is Heard? Civil Society Participation and Participatory Justice in DSA Systemic Risk Management* (https://dsa-observatory.eu/2026/02/17/who-speaks-and-who-is-heard).

Center for Countering Digital Hate (2022) *The Incelosphere: Exposing Pathways into Incel Communities and the Harms They Pose to Women and Children*. Center for Countering Digital Hate (https://counterhate.com/research/incelosphere).

Center for Democracy and Technology (2024) 'Trusted Flaggers in the DSA: Challenges and Opportunities'. Center for Democracy and Technology (https://cdt.org/insights/trusted-flaggers-in-the-dsa-challenges-and-opportunities).

Center for Intimacy Justice (2025) *The Digital Gag: Suppression of Sexual and Reproductive Health on Meta, Tiktok, Amazon, and Google* (www.intimacyjustice.org/report2025).

Collins, C.J. et al. (2025) 'A Review of Anti-Misogyny Interventions for Children and Adolescents: Recommendations for the Future'. *Journal of Gender Studies* 1–18 (https://doi.org/10.1080/09589236.2025.2552805).

D'Andrea, A., Fusacchia, G. and D'Ulizia, A. (2025) 'Policy Review: Countering Disinformation in the Digital Age. Policies and Initiatives to Safeguard Democracy in Europe'. *Information Polity* 30(1): 82–91 (https://doi.org/10.1177/15701255251318900).

DG Connect (2023) *Digital Services Act : Application of the Risk Management Framework to Russian Disinformation Campaigns*. Directorate-General for Communications Networks, Content and Technology, European Union (https://data.europa.eu/doi/10.2759/764631).

Di Meco, L. (2020) 'Online Threats to Women's Political Participation and The Need for a Multi-Stakeholder, Cohesive Approach to Address Them'. Expert paper. UN Women (www.unwomen.org/en/csw/csw65-2021/preparations/expert-group-meeting).

Diepeveen, S. (2022) *Hidden in Plain Sight: How the Infrastructure of Social Media Shapes Gender Norms*. ALIGN. (www.alignplatform.org/resources/report-hidden-in-plain-sight).

Ding, M.L., Suresh, H. and Venkatasubramanian, S. (2026) 'How to Stop Playing Whack-a-Mole: Mapping the Ecosystem of Technologies Facilitating AI-Generated Non-Consensual Intimate Images'. arXiv:2602.04759. Preprint, arXiv, 15 May (https://doi.org/10.48550/arXiv.2602.04759).

Douek, E. (2022) 'Content Moderation as Systems Thinking'. *Harv. L. Rev*. 136: 526.

DSA Civil Society Coordination Group (2025) *Initial Analysis on the First Round of Systemic Risk Reports Published under the DSA* (https://cdt.org/wp-content/uploads/2025/03/RA-Report-Assessment-Report.pdf).

DSA Civil Society Coordination Group (2026) 'Grok Floods X with Sexualized Images of Women and Children'. DSA Civil Society Coordination Group (https://counterhate.com/research/grok-floods-x-with-sexualized-images).

Dunn, S. (2020) 'Technology-Facilitated Gender-Based Violence: An Overview'. Centre for International Governance Innovation: Supporting a Safer Internet (https://digitalcommons.schulichlaw.dal.ca/scholarly_works/773).

Dunn, S. (2021) 'Is It Actually Violence? Framing Technology-Facilitated Abuse as Violence'. In *The Emerald International Handbook of Technology-Facilitated Violence and Abuse*. Emerald Publishing (https://papers.ssrn.com/abstract=3869759).

ECNL (2026) *Five Critical Lessons from Three Years of DSA Risk Assessments* (https://ecnl.org/publications/five-critical-lessons-three-years-dsa-risk-assessments).

Ekstrand, M.D., Riedl, J.T. and Konstan, J.A. (2011) 'Collaborative Filtering Recommender Systems'. *Foundations and Trends in Human-Computer Interaction* 4(2): 81–173 (https://doi.org/10.1561/1100000009).

Equality Now and Sexual Violence Prevention Association (2026) *Online Sexual Exploitation and Abuse in the United States: An Analysis of Policy Gaps, System Response and Prevention Mechanisms through Survivor-Lived Experiences*. Equality Now & Sexual Violence Prevention Association (https://equalitynow.org/wp-content/uploads/2026/01/Online-sexual-exploitation-and-abuse-in-the-United-States-An-analysis-of-policy-gaps-system-response-and-prevention-mechanisms-Full-report-English-2.pdf).

Eskens, S. (2025) 'Public or Private Communication? Categorising WhatsApp's and Telegram's Group Chat and Channel Functionalities under the DSA'. *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 16(3) (www.jipitec.eu/jipitec/article/view/446).

Estrela, E. (2013) *Report on Sexual and Reproductive Health and Rights*. A7-0426/2013. European Parliament (www.europarl.europa.eu/doceo/document/A-7-2013-0426_EN.html).

European Board for Digital Services (2025) *First Report of the European Board for Digital Services in Cooperation with the Commission Pursuant to Article 35(2) DSA on the Most Prominent and Recurrent Systemic Risks as Well as Mitigation Measures*. European Commission (https://home-affairs.ec.europa.eu/networks/eu-knowledge-hub-prevention-radicalisation/welcome-package/learning-resources/first-report-european-board-digital-services-cooperation-commission-pursuant-article-352-dsa-most_en).

European Commission (2018) 'Action Plan against Disinformation. European Commission Contribution to the European Council'. Communication. Brussels, 5 December (https://commission.europa.eu/publications/action-plan-disinformation-commission-contribution-european-council-13-14-december-2018_en).

European Commission (2026) 'The Impact of the Digital Services Act on Digital Platforms. Shaping Europe's Digital Future' (https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms).

European Commission (2026) 'Commission Investigates Grok and X's Recommender Systems under the Digital Services Act. Shaping Europe's Digital Future' (https://digital-strategy.ec.europa.eu/en/news/commission-investigates-grok-and-xs-recommender-systems-under-digital-services-act).

European Institute for Gender Equality (2022) Combating Cyber Violence against Women and Girls. EIGE (https://eige.europa.eu/publications-resources/publications/combating-cyber-violence-against-women-and-girls?language_content_entity=en).

Evans, S.K. et al. (2017) 'Explicating Affordances: A Conceptual Framework for Understanding Affordances in Communication Research' *Journal of Computer-Mediated Communication* 22(1): 35–52 (https://doi.org/10.1111/jcc4.12180).

Faith, B., Philip, R. and De Jesus Gomes, A. (2026) *Digital Backlash: Manifestations and Resistance*. Report. Institute of Development Studies (https://doi.org/10.19088/BACKLASH.2026.008).

Farrell, H. (2003) 'Constructing the International Foundations of E-Commerce: The EU–US Safe Harbor Arrangement'. *International Organization* 57 (2): 277–306 (https://doi.org/10.1017/S0020818303572022).

Farvid, P. et al. (2026) 'Tradwife Influencers and the Mainstreaming of Alt-Right Femininity: A Visual Motif Analysis of Gendered Extremism Online'. *Psychology of Women Quarterly* 50(2): 165–85 (https://doi.org/10.1177/03616843261442466).

Fisher, K., Rice, S. and Seidler, Z. (2025) Young Men's Health in a Digital World. Movember Institute of Men's Health (https://ex.movember.com/movember-institute/masculinities-report).

Fitzgerald, K. (2020) 'Mapping the manosphere: a social network analysis of the manosphere on Reddit'. Masters Thesis, Naval Postgraduate School (http://hdl.handle.net/10945/66637).

Freelon, D. et al. (2024) 'The Post-API Age of Social Media Data Access: Past, Present, and Future'. *ANNALS of the American Academy of Political and Social Science* 715(1): 16–37.

Gehrke, M. (2026) 'How Are Gender and Race Deployed in Disinformation? An Exploratory Study of Women Politicians'. *Journal of Gender Studies* 35(3): 648–64 (https://doi.org/https://doi.org/10.1080/09589236.2025.2600383).

Gehrke, M. and Amit-Danhi, E.R. (2025) 'Gendered Disinformation as Violence: A New Analytical Agenda'. *Harvard Kennedy School Misinformation Review* 6(3) (https://doi.org/https://doi.org/10.37016/mr-2020-177).

Gerrand, V. et al. (2025) 'Mapping the Neo-Manosphere(s): New Directions for Research'. *Men and Masculinities* 28(5): 443–64 (https://doi.org/https://doi.org/10.1177/1097184X251350277).

Gillespie, T. (2014) 'The Relevance of Algorithms'. In T. Gillespie, P.J. Boczkowski and K.A. Foot (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press (https://doi.org/10.7551/mitpress/9042.003.0013).

Gillespie, T. (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

Ging, D. (2019) 'Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere'. *Men and Masculinities* 22(4): 638–57. https://doi.org/10.1177/1097184X17706401.

Gorwa, R. (2019) 'What Is Platform Governance?' *Information, Communication & Society* 22(6): 854–71 (https://doi.org/https://doi.org/10.1080/1369118X.2019.1573914).

Gozzi, L. (2026) 'France Investigates Reappearance of Website Tied to Pelicot Crimes'. BBC news, 29 April (www.bbc.com/news/articles/com2ndp20420).

Griffin, R. (2025) 'Governing Platforms through Corporate Risk Management: The Politics of Systemic Risk in the Digital Services Act'. *European Law Open* 4(2): 223–53.

Haimson, O.L. et al. (2021) 'Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas'. *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 466:1-466:35 (https://doi.org/10.1145/3479610).

Haroon, M. et al. (2023) 'Auditing YouTube's Recommendation System for Ideologically Congenial, Extreme, and Problematic Recommendations'. *Proceedings of the National Academy of Sciences* 120(50): e2213020120.

Haukkamäki, V.E. (2025) 'Gender, Media, and the Politics of Visibility: The Case of Sanna Marin'. Thesis. University of Helsinki (https://helda.helsinki.fi/items/94fb7b20-2612-4072-b314-0f848feaef9c).

Hayes, C. et al. (2026) *Engaging Men, Ending Violence: Pathways to Safer Digital Futures*. UN Women and Equimundo (www.unwomen.org/en/digital-library/publications/2026/04/engaging-men-ending-violence-pathways-to-safer-digital-futures).

Hickey, D. et al. (2025) 'X under Musk's Leadership: Substantial Hate and No Reduction in Inauthentic Activity'. *PloS One* 20(2): e0313293.

Hinson, L. et al. (2018) *Technology-Facilitated GBV: What Is It, and How Do We Measure It?* Brief. International Center for Research on Women (ICRW) (www.icrw.org/publications/technology-facilitated-gender-based-violence-what-is-it-and-how-do-we-measure-it).

van Hoboken, J. et al. (eds) (2023) *Putting the DSA into Practice: Enforcement, Access to Justice and Global Implications*. Verfassungsbooks (https://doi.org/10.17176/20230208-093135-0).

Hörnle, J. (2026) 'Fighting the Beast of Image-Based Sexual Abuse. Part 1: The Criminal Law and Platform Regulation'. *International Journal of Law and Information Technology* 34: eaag006 (https://doi.org/10.1093/ijlit/eaag006).

Husovec, M. (2024) 'The DSA as a Co-Regulatory System'. In M. Husovec (ed.) *Principles of the Digital Services Act*. Oxford: Oxford University Press (https://doi.org/10.1093/law-ocl/9780192882455.003.0020).

Ignazzi, E., Maretti, M. and Fontanella, L. (2025) 'The Italian Manosphere: Composition, Structure, and Functions of a Digital Network'. *Social Inclusion* 13(0) (https://doi.org/10.17645/si.9341)

Institute for Strategic Dialogue (2020) *Hoodwinked: Coordinated Inauthentic Behaviour on Facebook* (www.isdglobal.org/publication/hoodwinked-coordinated-inauthentic-behaviour-on-facebook).

Institute for Strategic Dialogue (n.d.) Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere' (www.isdglobal.org/publication/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere).

Inter-Parliamentary Union (2018) *Sexism, Harassment and Violence against Women in Parliaments in Europe*. Issue Brief. Inter-Parliamentary Union (www.ipu.org/resources/publications/issue-briefs/2018-10/sexism-harassment-and-violence-against-women-in-parliaments-in-europe).

Jang, H. and Nanditha, N. (2025) 'Echo Chambers of Digital Harm: Insights Into Layered Affordances From India and South Korea'. *Social Media + Society* 11(4) (https://doi.org/10.1177/20563051251383528).

Kitchin, R. (2017) 'Thinking Critically about and Researching Algorithms'. *Information, Communication & Society* 20(1): 14–29 (https://doi.org/10.1080/1369118X.2016.1154087).

Klonick, K. (2017) 'The New Governors: The People, Rules, and Processes Governing Online Speech'. *Harvard Law Review* 131(6): 1598–670.

Kojah, S.A. et al. (2025) '"Dialing It Back:" Shadowbanning, Invisible Digital Labor, and How Marginalized Content Creators Attempt to Mitigate the Impacts of Opaque Platform Governance'. *Proceedings of the ACM on Human-Computer Interaction* 9(1) (https://doi.org/10.1145/3701191).

Korolczuk, E.; Graff , A. and Kantola, J. (2025) 'Gender danger. Mapping a decade of research on anti-gender politics'. *Journal of Gender Studies* (https://doi.org/10.1080/09589236.2025.2489584).

Le Monde (2026) 'Dominique Pelicot, Reconnu Coupable Dans l'affaire Des Viols de Mazan, Transféré à La Prison d'Ensisheim, Qui Accueille Des Condamnés à de Longues Peines'. 25 March (www.lemonde.fr/societe/article/2026/03/25/dominique-pelicot-reconnu-coupable-dans-l-affaire-des-viols-de-mazan-transfere-a-la-prison-d-ensisheim-qui-accueille-des-condamnes-a-de-longues-peines_6674188_3224.html).

Leerssen, P. (2023) 'An End to Shadow Banning? Transparency Rights in the Digital Services Act between Content Moderation and Curation'. *Computer Law & Security Review* 48. (https://doi.org/10.1016/j.clsr.2023.105790).

Leerssen, P. (2025) 'From Murdoch to Musk: Platform Ownership and the Political Economy of Online Content Governance'. *Platforms & Society* 2 (https://doi.org/10.1177/29768624251386260).

Leerssen, P. et al. (2025) *Pathways to Private Enforcement of the Digital Services Act*. DSA Observatory (https://dsa-observatory.eu/2025/06/05/report-pathways-to-private-enforcement-of-the-digital-services-act-dsa).

Leidig, E. (2023) *The Women of the Far Right. Social Media Influencers and Online Radicalization*. New York: Columbia University Press (www.ru.nl/en/services/sport-culture-and-recreation/radboud-reflects/news/the-women-of-the-far-right-social-media-influencers-and-online-radicalization-lecture-and-conversation-with-radicalization-expert-eviane-leidig-and-political-scientist-andrej-zaslove).

Lima-Strong, C. and Jehangir, R. (2025) 'How US Officials Are Pressuring Europe Over Its Platform Regulations'. TechPolicyPress (https://techpolicy.press/how-us-officials-are-pressuring-europe-over-its-platform-regulations).

Lott, K., Murumaa-Mengel, M. and Marling, R. (2025) 'Mainstreaming the Manosphere: Discourses of Contemporary Masculinity among Estonian Manfluencers'. *Humanities and Social Sciences Communications* 12(1): 602 (https://doi.org/10.1057/s41599-025-04903-y).

Magalhães, J.C., Keller, C.I. and Gorwa, R. (2026) 'The Great Sysop: Elon Musk, X, and the Emergence of Platform Illiberalism'. *New Media & Society* (https://journals.sagepub.com/doi/10.1177/14614448261424889).

Marwick, A.E. (2021) 'Morally Motivated Networked Harassment as Normative Reinforcement'. *Social Media+Society* 7(2).

McGinley, C. and Woods, L. (n.d.) I*mage-Based Sexual Abuse, Pornography Platforms and the Digital Services Act*. Expert opinion, University of Durham and University of Essex (https://hateaid.org/wp-content/uploads/2022/04/ImageBasedAbuse-and-DSA-Expert-Opinion-McGlynn-and-Woods-17-Jan-2022.pdf).

Michalko, J. (2026) *Online Violence against Women in Politics: What Shapes Political Party Responses to Technology-Facilitated Gender-Based Violence?* ALIGN (www.alignplatform.org/resources/report-online-violence-against-women-in-politics).

Milli, S. et al. (2025) 'Engagement, User Satisfaction, and the Amplification of Divisive Content on Social Media'. *PNAS Nexus* 4(3) (https://doi.org/10.1093/pnasnexus/pgaf062).

Observacom (2025) *El shadow banning: la censura sutil y oculta de las grandes plataformas digitales*. Observacom (www.observacom.org/el-shadow-banning-la-censura-sutil-y-oculta-de-las-grandes-plataformas-digitales).

OECD (2022) *OECD Framework for the Classification of AI Systems*. OECD Publishing (https://doi.org/10.1787/cb6d9eca-en).

Oz, M., Shahin, S. and Greeves, S.B. (2024) 'Platform Affordances and Spiral of Silence: How Perceived Differences between Facebook and Twitter Influence Opinion Expression Online'. T*echnology in Society* 76 (https://doi.org/10.1016/j.techsoc.2023.102431).

Palumbo, A. (2026) 'Systemic Risk Management and the Constitutional Limits of Delegating Political Discretion: An Analysis of the DSA and the AI Act'. *European Journal of Risk Regulation* 17(2): 454–75 (https://doi.org/10.1017/err.2025.10069).

Papapietro, G. and Duprat-Macabies, A. (2024) 'Online Gender-Based Violence in the EU: What Now?' Center for Democracy and Technology (https://cdt.org/insights/online-gender-based-violence-in-the-eu-what-now).

Posetti, J. and Shabbir, N. (eds) (2022) *The Chilling: A Global Study On Online Violence Against Women Journalists*. International Center for Journalists (ICFJ) (www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists).

Rajani, N. (2022) '"I Bet You Don't Get What We Get": An Intersectional Analysis of Technology-Facilitated Violence Experienced by Racialized Women Anti- Violence Online Activists in Canada'. *Canadian Journal of Law and Technology* 19(2) (https://digitalcommons.schulichlaw.dal.ca/cjlt/vol19/iss2/2).

Rathje, S., Van Bavel, J. and Van Der Linden, S. (2021) 'Out-Group Animosity Drives Engagement on Social Media'. *Proceedings of the National Academy of Sciences* 118(26): e2024292118.

Regehr, K. et al. (2024) 'Safer Scrolling'. Association of School and College Leaders.

Reichelmann, A., Hawdon, J., Costello, M., et al. (2021). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior* 42(9): 1100–1111 (https://doi.org/10.1080/01639625.2020.1722337).

Rigotti, C., Döring, N. and Rossoni, I. (2026) 'Very Large, Largely Silent: How Three Pornography-Hosting Platforms Assess Their Governance of Sexual Content under the Digital Services Act'. *Porn Studies* (https://doi.org/10.1080/23268743.2026.2692431).

Santos, F.G. and Geva, D. (2022) 'Populist Strategy in the European Parliament: How the Anti-Gender Movement Sabotaged Deliberation about Sexual Health and Reproductive Rights'. *European Journal of Cultural and Political Sociology* 9(4): 475–501 (https://doi.org/10.1080/23254823.2022.2113417).

Scheele, C. (2024) 'The Power of Digital Platforms in Standing up for the Safety of Women and Girls Online. European Institute for Gender Equality'. EIGE. 26 November (https://eige.europa.eu/newsroom/director-corner/power-digital-platforms-standing-safety-women-and-girls-online).

Schnidrig, D. (2026) *Bridging the Gap: Addressing Technology-Facilitated Gender-Based Violence in Global AI Governance*. Association for Progressive Communications (www.apc.org/en/pubs/bridging-gap-addressing-technology-facilitated-gender-based-violence-global-ai-governance).

Schwartz, B. and Neff, G. (2019) 'The Gendered Affordances of Craigslist "New-in-Town Girls Wanted" Ads'. *New Media & Society* 21(11–12): 2404–21 (https://doi.org/10.1177/1461444819849897).

Science, Innovation and Technology Committee (n.d.) *Social Media, Misinformation and Harmful Algorithms* (https://publications.parliament.uk/pa/cm5901/cmselect/cmsctech/441/report.html).

Semenzin, S. and Bainotti, L. (2020) 'The Use of Telegram for Non-Consensual Dissemination of Intimate Images: Gendered Affordances and the Construction of Masculinities'. *Social Media+Society* 6(4) (https://doi.org/10.1177/2056305120984453).

Sinders, C. (2022) *Designing against Dark Patterns*. GMF policy paper (www.gmfus.org/news/designing-against-dark-patterns).

Singh, S. (2019) *Everything in Moderation: How Platforms Use AI to Moderate User-Generated Content*. New America (www.newamerica.org/insights/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content).

Southerton, C. et al. (2021) 'Restricted Modes: Social Media, Content Classification and LGBTQ Sexual Citizenship'. *New Media & Society* 23(5): 920–38 (https://doi.org/10.1177/1461444820904362).

Strowel, A. and De Meyere, J. (2023) 'The Digital Services Act: Transparency as an Efficient Tool to Curb the Spread of Disinformation on Online Platforms?' *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law* 14(1): 66–83 (https://doi.org/https://nbn-resolving.de/urn:nbn:de:0009-29-57089).

Suler, J. (2004) 'The Online Disinhibition Effect'. Cyberpsychology, Behavior, and Social *Networking* 7(3): 321–26 (https://doi.org/10.1089/1094931041291295).

Suzor, N. et al. (2019) 'Human Rights by Design: The Responsibilities of Social Media Platforms to Address Gender-Based Violence Online'. *Policy & Internet* 11(1): 84–103 (https://doi.org/10.1002/poi3.185).

Sykes, S. and Hopner, V. (2024) 'Tradwives: Right-Wing Social Media Influencers'. *Journal of Contemporary Ethnography* 53(4): 453–87 (https://doi.org/10.1177/08912416241246273).

Tabbush, C. (2025) Understanding backlash against gender equality: Evidence, trends and policy responses. UN Women and UNRISD (https://knowledge.unwomen.org/en/digital-library/publications/2025/09/understanding-backlash-against-gender-equality-evidence-trends-and-policy-responses).

Tech.Eu (2025) 'Startups Take on Big Tech: Women's Health Companies File EU Complaints over Digital Censorship'. 10 March (https://tech.eu/2025/03/10/startups-take-on-big-tech-womens-health-companies-file-eu-complaints-over-digital-censorship).

Tyabashe-Phume, B.P. et al. (2025) 'Scoping Review on Technology-Facilitated Gender-Based Violence against Women with Disabilities and LGBTQI+ Persons in Low- and Middle-Income Countries'. *African Journal of Disability* 14 (https://doi.org/10.4102/ajod.v14i0.1810).

UN Women (2023) *Technology-Facilitated Violence against Women: Taking Stock of Evidence and Data Collection*. UN Women (www.unwomen.org/en/digital-library/publications/2023/04/technology-facilitated-violence-against-women-taking-stock-of-evidence-and-data-collection).

UN Women (2025) 'FAQs: Digital Abuse, Trolling, Stalking, and Other Forms of Technology-Facilitated Violence against Women and Girls'. 13 November (www.unwomen.org/en/articles/faqs/digital-abuse-trolling-stalking-and-other-forms-of-technology-facilitated-violence-against-women).

UN Women Regional Office for Europe and Central Asia (2023) *The Dark Side of Digitalization: Technology-Facilitated Violence against Women in Eastern Europe and Central Asia*. UN Women Regional Office for Europe and Central Asia (https://eca.unwomen.org/en/digital-library/publications/2023/11/the-dark-side-of-digitalization-technology-facilitated-violence-against-women-in-eastern-europe-and-central-asia).

Uwazuruike, A. (2026) 'Censorship of Women's Health and Wellbeing Content Online'. House of Commons Library (https://commonslibrary.parliament.uk/research-briefings/cdp-2026-0052).

Van der Vlist, Fernando, N. and Helmond, A. (n.d.) 'How Partners Mediate Platform Power: Mapping Business and Data Partnerships in the Social Media Ecosystem'. *Big Data & Society* 8(1) (https://journals.sagepub.com/doi/10.1177/20539517211025061).

Women's Aid (2026) 'Submission to the European Commission and the Board of the Digital Services Coordinators on the Article 35(2) DSA Report on Systemic Risks and Their Mitigation' (www.womensaid.ie/app/uploads/2026/04/Submission-to-the-European-Commission-and-the-Board-of-the-Digital-Services-Coordinators.pdf).

Appendix 1 Key Informant Interviews

1. Policy Officer at a European rights network
2. Director at democracy and technology CSO
3. Policy and research staff member of a gender equality CSO
4. Journalist and consultant on intersection of journalism, technology and human rights
5. Cybersecurity academic expert
6. Internet law academic expert
7. EU Senior Policy Officer at digital rights CSO
8. Board member and founder of AI system-auditing CSO
9. Law academic expert
10. EU digital policy negotiator
11. Digital Policy Analyst for a European think tank
12. Researcher on law and theory of technology
13. Postdoctoral researcher on social media regulation
14. Project Leader at a development agency
15. Scholar and activist with a focus on gender-related violence, digital rights, and emerging technologies
16. Executive Director of a digital rights CSO
17. Director and Human Rights Lawyer at a global NGO

ODI Europe

ODI Europe brings research, ideas and dialogue with the wider world into the European policy debate. We address critical priorities while staying attuned to European policies and processes. We nurture global research excellence across our thematic priorities to inspire policy changes for a fairer and more sustainable world. We are based in Brussels and are part of the ODI Global network.

ODI Europe
35 Square de Meeûs
1000 Brussels
Belgium

EU transparency register:
051479091677-17

odieurope@odi.org

Website: odi.org
LinkedIn: ODI Global and ODI Europe
BlueSky: @odi.global
Threads: @odi.global
Instagram: @odi.global