WARNING: This consultation response contains language and/or material that may be distressing

Ofcom Draft VAWG Guidance Consultation

End Violence Against Women Coalition Response
Dated 23.05.25

End Violence Against Women Coalition (EVAW) is a policy and campaigns charity, with a membership of 160+ specialist VAWG services, NGOs, academics and survivors working to end violence against women and girls (VAWG) in all its forms.

EVAW was originally set up in 2005 by a collection of women's specialist VAWG services with the objective of obtaining a cross-governmental VAWG strategy. Since becoming a registered charity, the charity occupies a distinct role due to its ability to represent and support our members, convene and build consensus within the VAWG sector and provide an independent expert voice to government and other stakeholders. In response to Ofcom's consultation on the document 'A safer life online for women and girls - practical guidance for tech companies', EVAW coordinated a joint response on behalf of the Online VAWG Network in the form of an oral transcript, as well as a joint briefing on behalf of the VAWG sector.

We know that online abuse is widespread, and disproportionately impacts women and girls. 77% of girls and young women aged 7-21 have experienced online harm in the last year[1], whilst 1 in 14 adults have experienced threats to share their intimate images without their consent[2] and the Revenge Porn Helpline dealt with 60,000 cases since it was founded 10 years ago, with reports rising on average 57% every year[3]. A report published by Ofcom stated that two thirds of UK internet users over the age of 16 came across a deepfake image in the first half of 2024[4], indicating that fake content already exists in abundance online.

Indeed, the Revenge Porn Helpline has seen a 400% increase in cases involving synthetic sexual content, otherwise known as deepfakes, between 2017-2024. This is all whilst operating in a landscape where protections for users are being eroded, and with a general trend of tech companies delivering the bare minimum in terms of online safety, where we are already seeing a concerning rollback of protections for users on major social media platforms such as Meta.

---

1 https://www.girlguiding.org.uk/girls-making-change/girls-attitudes-survey/#:~:text=This%20survey%20asked%20over%202,000%20girls%20and

2 https://refuge.org.uk/news/intimate-image-abuse-despite-increased-reports-to-the-police-charging-rates-remain-low/

3 https://www.theguardian.com/society/2025/feb/22/revenge-porn-abusers-devices-illicit-images

4 https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2024/online-nation-2024-report.pdf?v=386238

WARNING: This consultation response contains language and/or material that may be distressing

In recent years, we have also seen the growth of online misogynist influencers who promote violence against women and girls and have had a real impact on boys and men's attitudes and behaviour[5]. Our members continue to see abusers utilise developments in technology[6] to further abuse, with tech companies promoting and profiting from such harmful content.[7]

Ofcom, whilst hamstrung by their inability to make this an enforceable code of practice, could go further when interpreting the platforms duties through both codes in relation to women and girls, as well as providing more ambitious suggestions for platforms in the good practice section. Our response details numerous ways in which this could be achieved.

Question 1: Do you have any comments on our proposed approach to 'content and activity' which 'disproportionately affects women and girls'?

Scope of VAWG Covered

Ofcom's draft Guidance focuses on content and activity that disproportionately affects women and girls, proposing four key harm areas: Online misogyny, Pile-ons, Domestic abuse, and Image-based sexual abuse. These are referred to collectively as online gender-based harms, which are understood as systemic and intersectional, driven by long standing forms of misogyny and sexism that overlap with factors such as age, race, ethnicity, and disability. While the guidance acknowledges that other harms like child sexual exploitation and abuse (CSEA) also disproportionately impact girls, it notes that these are primarily addressed under separate sections of the Act.

However the Ofcom draft guidance has notable gaps in addressing stalking, domestic abuse, and misogynoir, each of which is inadequately covered in ways that undermine the protection of vulnerable groups and the guidance's overall effectiveness.
Stalking is missing from the four key harm areas and as such is presented in a limited manner, with insufficient recognition of its unique nature as a course of conduct crime. Stalking involves repeated, unwanted behaviours that form a pattern over time, making the course of conduct itself the crime. Despite this, platforms frequently fail to acknowledge these patterns, focusing instead on whether individual incidents meet harm thresholds. This lack of awareness impedes platforms' ability to take effective action.

5 https://www.bbc.co.uk/news/articles/cne4vw1x83po

6 https://www.theguardian.com/technology/2025/feb/01/stalking-ai-chatbot-impersonator

7 https://www.bbc.co.uk/future/article/20250207-how-google-amazon-and-microsoft-funnelled-ad-money-to-a-site-hosting-child-abuse-images

Additionally, the guidance does not adequately address the interconnected nature of online and offline stalking. Behaviours that escalate online often have real-world implications, increasing the risk to victims, yet this critical connection is overlooked. Furthermore, the guidance lacks recommendations for sharing data across platforms, which could help identify patterns of stalking that cross platform boundaries, further disadvantaging victims trying to prove a pattern of abuse.

Likewise for domestic abuse, the guidance falls short by narrowly framing it as an issue affecting adult women, neglecting the experiences of girls under the age of 16. This omission perpetuates misconceptions and ignores the reality that young girls also face technology-facilitated domestic abuse. Survivors also face practical challenges in accessing evidence of abuse, such as logs of harmful interactions or platform responses, which are often critical for legal proceedings. The absence of clear guidance for platforms to facilitate such access leaves survivors without the support they need to seek justice.

We welcome the inclusion of intimate image abuse as a specific form of harm but recommend that it is expanded to include 'semen images', where a woman's image is posted online then using generative AI semen is added onto the image[8].

Misogynoir, the intersection of racism and misogyny, is acknowledged in the guidance but not thoroughly addressed. The definition of misogynoir and how platforms should detect and manage it remain vague. While a case study on automated detection of misogynoir content is included, it does not provide practical recommendations for implementation, leaving platforms without clear directives. Additionally, the guidance fails to account for the disproportionate impact content moderation systems have on Black women, particularly when they engage in counter-speech or share personal experiences of discrimination and harm. This oversight can lead to the silencing of Black women's voices, further marginalising them in online spaces.

Algorithm Design

We particularly welcome the recommendation in Action 6 to train *"content recommendation algorithms to be gender-sensitive could include: Auditing and evaluating recommender algorithms and other AI systems to assess whether they promote online misogyny, as well as evaluating gender bias in recommendations and retraining algorithms either after revising existing datasets*" (page 41). However the

8 https://researchbriefings.files.parliament.uk/documents/LLN-2024-0070/2024-0070-Non-Consensual-Sexually-Explicit-Images-and-Videos-(Offences)-Bill-[HL]-LARGE.pdf

harm caused by algorithm design must be threaded through the guidance by recognising it as a standalone form of harm.

The inclusion of algorithmic design as a specific form of harm is essential to recognise the way in which algorithms shape online environments and can perpetuate or amplify misogynistic content and harmful behaviour. Recommendation systems often prioritise content which includes misogynistic, abusive, or threatening material targeting women and girls. By failing to address how these algorithms function, platforms risk creating environments that not only tolerate but amplify and profit from such harm.

Algorithms can also play a role in facilitating targeted harassment. Predictive tools and pattern recognition technologies may expose women to further abuse by prioritising comments or posts from aggressive users if these interactions drive engagement. Similarly, algorithmic bias can perpetuate systemic gender and racial inequality. For example, algorithms trained on biased datasets may undervalue women's contributions or disproportionately flag content discussing issues like misogyny or sexual violence as inappropriate while under-policing actual harmful behaviour or content promoting misogynistic ideology. Algorithmic design can be particularly harmful to women from marginalised groups, such as Black and minoritised women, the LGBTQ+ community, or those with disabilities, who may face compounded harms due to intersecting vulnerabilities.

Addressing these concerns explicitly in the VAWG guidance ensures that platforms are held accountable for the societal impact of their technologies and underscores the need for algorithmic transparency, regular auditing to identify harmful patterns, and the integration of ethical AI practices that promote gender transformative content and diversity.

Pornography

The guidance must comprehensively address the role of pornography in perpetuating harmful norms and behaviours against women and girls by expanding its scope and focus. The current emphasis on child protection should be broadened to include the cultural impacts of pornography, such as its role in normalising violence and promoting harmful stereotypes. This includes addressing non-illegal but harmful content like incest or strangulation pornography, which is linked to increased instances of physical harm and risky behaviours. Moreover, platforms' algorithms that promote harmful pornographic content need greater scrutiny. These systems often amplify misogynistic material, further embedding harmful norms within digital spaces. In addition, technologies such as nudification and deepfake apps should be explicitly analysed, with detailed recommendations on mitigating their accessibility and impact.

WARNING: This consultation response contains language and/or material that may be distressing

To ensure effective oversight, the guidance should advocate for robust age verification systems and mandate that platforms incorporate pornography-related risks into their broader risk assessments. Enhanced measures for accountability, such as preventing the circulation of harmful content and ensuring compliance through transparency reporting are essential.

Minimum steps

Whilst the guidance introduces some positive measures that this guidance takes to address VAWG, we're concerned that measures that will have the most impact are framed within the good practice section of the guidance, which relies on the good will of platforms to implement. The foundational steps set out in the guidance are those which platforms must take in order to comply with the Illegal Harms Code and the Children's Code, and are legally required of platforms through the new regulatory system. However the framing of these steps as 'foundational' could be interpreted as aspirational for tech platforms, with a lack of emphasis on their new duties.

We would advocate that instead, 'Foundational steps' is changed to 'minimum steps', which clearly communicates that these steps represent a baseline standard that platforms must meet to be compliant with their duties as set out in the Children's Code and Illegal Harms Code. This distinction is crucial to avoid misinterpretation by platforms that these measures are in any way optional or flexible under the current regulatory framework. Minimum steps would better reflect that platforms must go further to properly ensure women and girls safety is prioritised.

We would also point out that whilst the guidance provides some detailed case studies relating to intimate image abuse and stalking which outline some positive steps, they are placed within the good practice section. As both stalking and intimate image abuse are listed as priority offences, we would argue that they should be placed in the minimum steps section to reflect the duties that platforms have to protect their users from those offences.

Intersectional Framework

Violence, abuse and discrimination online are shaped by overlapping and compounding systems of oppression, such as racism, sexism, classism, ableism, homophobia or transphobia. For women and girls, particularly those from Black and minoritised communities, this means online harms are not simply gendered but also racialised and often invisibilised within current moderation and safety policies. In Glitch's work on misogynoir, they highlight the unique forms of racialised misogyny directed at Black

women, which are frequently unrecognised in content moderation and algorithmic oversight[9].

Applying an intersectional framework to tech platforms would involve fundamentally rethinking safety by design principles to move beyond generic harm reduction to actively embedding protections that are responsive to the specific needs of the most marginalised users. This includes building moderation systems that detect and respond to intersectional forms of abuse, involving a diverse group of women in the product design and internal policy making to ensure that reporting mechanisms and other safety features are informed by nuanced, identity-specific harms.

An intersectionality framework would also acknowledge how algorithms both reproduce racial and gender bias, and actively promote content that is racist. There must be greater transparency and accountability in algorithmic decision-making, particularly where content recommendations disproportionately affect Black and minoritised women.

Platforms should take deliberate steps to ensure that Black and minoritised women and girls have an equal voice, visibility and influence on platforms. This includes protecting users from online harassment, amplifying diverse perspectives and fostering inclusive online communities. Tech providers must recognise that when the ecosystem of their platform contributes to the perpetration of abuse, black and minoritised communities that are disproportionately subject to online abuse may come off their platform altogether, hindering their freedom of speech and contributing to a lack of representation and visibility on the platform. An intersectional framework requires platforms to design out any features that could result in harm for the most at risk users, prioritising their digital rights.

Harm Prevention

We welcome the focus on features and functionalities, and how they can facilitate abuse, deterring perpetrators and ensuring platforms are safe by design to prevent harm occurring in the first place. This indicates a more proactive approach to identifying where design features of a platform could be used to perpetrate abuse.

However whilst Chapter 4 is called ‘Harm Prevention’, Ofcom’s must go further to set out how tech platforms can truly ensure they are safe by design, taking a proactive approach to address the root causes of VAWG and systems that enable harmful behaviours, aiming to stop online abuse before it occurs. While responding to harm is

[9] https://drive.google.com/file/d/13x4P2ANjhJkx3Qodd0E3owrXmkE-ZTVJ/view

important, focusing on prevention helps ensure that the problem doesn't escalate or repeat in the first place, fostering a safer environment from the outset.

Primary prevention efforts must involve ensuring that algorithms that shape user experience do not contribute to the spread of harmful content online. This involve deprioritising content that is misogynistic in nature or reinforces sexist stereotypes or promotes harmful gender norms, while actively promoting gender-transformative content that challenges these biases. Such algorithmic reform requires a conscious shift in design practices currently being used by tech platforms, as set out above, to ensure that algorithms are not amplifying misogynistic and harmful content.

Safety by design involves considering safety from the initial product specification and design, through deployment, maintenance, and retirement. While the guidance acknowledges this temporal scope, other aspects like technical scope are not covered robustly enough, including looking beyond just takedown measures to account creation, distribution, and user engagement.

Safety by design is poorly defined in point 1.13 and it is therefore unclear throughout the document. The priority should be to design out risks before relying on mitigation and remediation. We would suggest a clearer link to the 9 proposed actions set out in the guidance when defining safety by and explained that in order for a platform to be defined as ‘safe by design’ all 9 actions must be taken.

Media Literacy

We welcome the systemic understanding of VAWG prevention laid out in chapter 4 of the guidance, and the focus on media literacy and education. Whilst there are some positive measures highlighted, the guidance defers to the strategy set out in the media literacy guidance that Ofcom published recently. Better media literacy is needed for all users, not just children. Whilst we welcomed Ofcom’s commitment to publish Media Literacy guidance, it must be more ambitious to achieve long-term attitudinal change and prevention of VAWG.

EVAW and Glitch wrote to Ofcom to outline our concerns regarding the media literacy guidance, which included:

- The media literacy strategy lack specificity and enforceability, particularly in holding tech companies accountable.
- Media literacy should be treated as an essential pillar of harm prevention, not just harm response, with clearly defined outcomes, robust evaluation and adequate resourcing to ensure long-term impact.

WARNING: This consultation response contains language and/or material that may be distressing

- Ofcom must embed an intersectional lens throughout the strategy, centering Black and minoritised women and girls, and recognising how systemic inequalities shape online harm.
- Media literacy efforts must focus on challenging perpetrator behaviour and platform design that fosters harm, not placing responsibility solely on users or survivors to protect themselves.
- Civil society organisations must be properly funded and not expected to participate in unpaid convenings or unfunded programme delivery - groups must be engaged and supported meaningfully.
- A single media literacy week is insufficient. Ofcom should develop a systemic, year-round approach which is aligned with government policy and focused on long-term, structural change.

We believe the overall ambition of Ofcom's role in media literacy should be to become a world leader in preventing and reducing online harms. In particular, this means Ofcom taking an ambitious and proactive strategy to hold media literacy at the heart of its work - building a holistic approach which is consistent across the Online Safety Act (OSA). This importantly includes building media literacy into OSA codes of practice. As it currently stands, we are disappointed at what seems to be lack of ambition at the heart of the three year strategy, and the absence of detail attached to it. We hope that this approach does not reflect Ofcom's prioritisation and investment in media literacy more widely. It is a fundamental part of tackling all forms of online harm and will have a clear impact on Ofcom's OSA enforcement work if prevention strategies are properly implemented with and by tech companies, as well as by civil society and Ofcom.

We are disappointed at the absence of detail in Ofcom's media literacy strategy as to what will be delivered and the lack of clarity, particularly around its measures of success. Given the focus placed on evaluation within the strategy we would have expected more attention to how the implementation of the strategy itself will be measured and evaluated. The strategy also contains no detail as to the extent of resource attached to it. We would be grateful for this information to be shared, as it is such a significant consideration in its execution and the level of impact that will result.

Question 2: Do you have any comments on the nine proposed actions? Please provide evidence to support your answer.

<u>Action 1: Taking Responsibility</u>

We broadly welcome the good practice steps proposed, particularly the recommendation for external oversight, as illustrated in the accompanying case study,

which improves accountability. However, the inclusion of media literacy by design in this section appears misaligned with the concept of taking responsibility and could be better placed in Chapter 4, with more detail about how this might look in practice.

It is also positive to see guidance to consult with subject experts, such as representatives from the VAWG sector, which we have long pushed for, but clarification is needed on how this will be resourced. We recommend that remuneration for experts and those with lived experience is included and standardised to ensure there is no further pressure on individuals and organisations with already stretched capacity.

A recurring issue across this section is the conflation of workplace systems with policies that alter platform design. There is a need for greater clarity to distinguish between internal workplace practices and external platform safety responsibilities.

The foundational steps outline that terms and statements have clear and accessible provisions on how users are protected from illegal content (including illegal harms that disproportionately affect women and girls, such as stalking and intimate image abuse) as well as content harmful to children. Whilst we welcome this inclusion, and agree that this is the bare minimum expected of platforms, we would argue that there needs to be a requirement put on services to ensure there is no rollback on what's in the terms of service[10].

We particularly welcome that Ofcom sets out that these actions should be an ongoing exercise where providers can continually assess and improve the experiences of women and girls on their service, however would reinforce that this is something tech platforms *must* be doing.

Action 2: Conduct Risk Assessments that Focus on Harms to Women and Girls

It's critical that platforms take a proactive approach to safeguarding, embedding it in design rather than reacting post-harm. Section 3.17 is strong, recognising priority illegal offences that disproportionately impact women, such as intimate image abuse, and the value of tracking a platforms' track record of dealing with complaints in risk assessments. However there must be more recognition of how the ecosystem of a platform and the algorithmic design facilitates the priority illegal content set out in the codes, and how this should be assessed.

Within this section we note a deference to safety measures set out in the codes of practice, however the 'triple shield' user empowerment model has been criticised for

10 https://www.onlinesafetyact.net/analysis/meta-s-rollback-of-protections-for-users-why-the-uk-government-needs-to-act-and-fast/

putting the onus on individuals to keep themselves safe on platforms rather, as well as the failure to apply these measures as on by default.

We particularly welcome the interpretation of platform duties through the codes to use user base demographics to show that online harassment disproportionately affects women and girls and monitor how the functionalities of the service and business model contribute to harassment. This is an important recognition of the role platforms can and should play to collect data that creates a picture of abuse happening on their platforms, which can be used to improve the sites, and should be published for transparency.

Whilst the VAWG Guidance emphasises the importance of risk assessments, we would advocate for more specific requirements on platforms to publish risk assessments. There is a necessity for a specific framework for equalities analysis of risk against women and girls with overlapping protected characteristics, considering intersectional inequalities throughout the risk assessment. This could be better highlighted through the case study provided. Risk assessments should be accompanied by a mitigation plan that addresses the issues raised by risk assessments as well as long term evaluation of safety measures that are introduced.

It is increasingly common for regulated services to contract out parts of their business function, including contracting out their safety measures, and the impact of that could be felt by people who are customers or people who work for the supplier such as moderators with poor working conditions. It will also mean that the VAWG guidance will not be applicable to areas of work that have been contracted by other organisations, meaning there is a significant regulatory gap.

Risk assessments should therefore include an assessment arising from business relationships. It is not enough for service providers to just draw supplier's attention to the Terms of Service and Community Standards, but should also include provision in any contractual documents, and ensure that those provisions are enforced where necessary

Regulated services which outsource any part of their business, including moderation of content, applications, GIFs, images, or any other content or tools, as well as safety tech, should ensure the vendor adheres to the social media provider's Terms of Service and Community Standards, and where necessary take action to enforce those standards, and that they have employee and mental health protection policies in place that adhere at least to the same standard. Outsourced user safety tools must also be fully compliant with the platforms' duties through the Codes as well as the recommendations made in this guidance. This recommendation is important in ensuring that regulated services do

not seek to avoid responsibility for VAWG through outsourcing or ignoring the human rights and harms risks arising from it.

Lastly, it is good to see a recognition that survivors' voices and needs should be included in section 3.19, however it's also important that this work is not tokenistic or extractive and leads to meaningful change to platform design if they're going to be included in the process. Likewise, as this guidance is also for combatting harm against girls, young people's voices should also be included in the design process to ensure their needs are met. There is a recommendation that tech platforms use external assessors, but we would welcome clarity on who this would be.

Action 3: Be transparent about women and girls' safety online

We strongly support Ofcom's use of their transparency powers to inform regulation and service provider practice. The use of Ofcom's transparency powers to analyse trends over time will be invaluable to policy makers to understand how platforms are complying with measures in the Act, as well as how emerging threats and trends will impact users. This information will be key for specialist service providers who support victims of VAWG. As such, we urge Ofcom to make full use of these powers and consult with civil society about the kind of data requests that would most service the needs to better support victims.

We particularly welcome the acknowledgment that transparency reporting should help lead providers of categorised services to take measures to reduce harms stemming from their activities.

It is also positive to see information and best practice sharing recommended within the introduction to this section, however it is unclear where this sits in the foundational practice or good practice set of recommendations. Interventions improving user safety or deterring online gender-based harms should be shared and adopted more widely by other platforms, which could be facilitated by Ofcom.

This section is an appropriate place for Ofcom to set out how they will use their transparency powers to track and measure take up of the VAWG guidance, which has not included in this document despite hearing from Ofcom at stakeholders meetings that this is what they intend to do. We are yet to hear how take up of the guidance will be monitored on non-categorised services.

We particularly welcome inclusion of racially disaggregated data in the good practice section, which is incredibly lacking currently, however we would push for further clarity

from Ofcom as well as specific examples of the kind of data they'll be requesting that could inform future platforms design and improve women and girls' safety.

Action 4: Conduct usability assessments and product testing

Usability testing must anticipate misuse by perpetrators. Reference 4.18s should more explicitly discuss how product design can enable or deter offences like intimate image abuse.

We welcome recommendations for platforms to partner with subject matter experts and use red teaming, but would encourage that these evaluations consider lived experiences of VAWG survivors.

There are opportunities to expand on the media literacy design principles included here, particularly incorporating the feedback ourselves and Glitch have raised which is detailed in our answer to question 1.

Action 5: Set safer defaults

We strongly support the proposal that safety settings should be on by default, but this must go beyond just children and young people. Limiting the strongest protections to cover minors alone will undermine protections for adult users, particularly those who are vulnerable, including young women aged 18+. Furthermore, the minimisation of the platform's responsibility in stating that adult users are "more capable" of managing safety (4.29) is highly problematic. 18 is a critical time for young people navigating the online world, particularly as we know that young women are disproportionately experiencing online VAWG[11], but this is the point where protections will be lost.

It is positive to see some inclusion of strong interaction and privacy default measures which will mitigate strangers being able to access information about users that could be used to perpetrate VAWG, however this should be extended to include default content filters on harmful content. We note that recommendations like 'safe search' only applies to larger platforms, however, as noted previously, smaller platforms often facilitate the highest harm and should therefore be considered in this guidance.

Platforms make design choices about whether to provide these tools and how easy they are to find and use (including providing instructions and examples in multiple languages). Given the tendency of users not to change the original settings, providers

[11] https://oro.open.ac.uk/96398/1/OVAW%204N%20full%20report%20%28March%202024%29.pdf

should have maximum safety settings within the platform as default (even if users can then change these settings.

Action 6: Reduce circulation of harmful content

This section is particularly comprehensive, including references to the 'safety work' burden on users and the need for platforms to reduce this by mitigating against the risk of harmful online content. In the foundational steps section we welcome Ofcom's recommendation that pornography, as well as misogynistic content, must not appear on children's feeds at all, however we would argue that recommender systems should be expanded to include all users in relation to harmful content unless they have actively searched for it.

It is also positive to see Child Sexual Abuse Material (CSAM) warnings when users are searching for CSAM content that flag information and support. We would encourage this to be extended to any searches for illegal content given that platforms have a duty in the Illegal Harms Code of Practice to tackle VAWG crimes such as intimate image abuse and stalking occuring on their platform. We also recommend that 'harmful content' referenced when signposting children to support also include extreme misogynistic content that may not be considered illegal. Misogynistic content is increasingly being served to young social media users and normalises harmful gender norms and gender inequality which underpins VAWG[12].

In relation to user verification, the guidance touches on privacy concerns with this policy but fails to give details on what they are. This should be expanded in order to be clear about the balance tech platforms can and should strike.

Recommendations in the good practice for deterrence messaging to stop users posting harmful content is a welcome addition, however we would urge Ofcom to go further to include abusive messages to avoid only focusing on harmful content rather than mitigating against abusive user behaviour.

We particularly welcome recommendations to de-monetise user-generated content which promotes online gender-based harms but is not clearly illegal to prevent it from earning advertising income. The section goes on to clarify that material which includes survivors talking about the harm they've faced must not be taken down mistakenly, an important nuance to include. The overall de-monetising approach, which is reinforced again in case study 13, is vital to combatting online gender-based harms and misogynistic content to ensure that users and platforms are not profiting from the abuse

[12] https://www.ucl.ac.uk/news/2024/feb/social-media-algorithms-amplify-misogynistic-content-teens

of women and girls. We also add the importance of ensuring that educational materials intended to address harms of misogyny online and abuse are not also mistakenly removed.

Action 7: Give users better control over their experiences

The foundational steps set out in this section are limited in scope, as most platforms already offer blocking and muting of other users, particularly larger platforms that these measures apply to. Whilst blocking can provide users the control to limit harmful content and messages from other users, it is important to recognise that in some cases, such as for victims of stalking or domestic abuse, blocking the user could risk missing the escalation of abuse which would inform both the individual and supportive agencies response. We urge Ofcom to be more ambitious in their interpretation of platforms' duties through the codes, going beyond control tools that put the onus on individuals to keep themselves safe from abuse.

The inclusion of support materials for child users when implementing control settings is welcome, however we encourage Ofcom to outline here the role civil society could play to inform them. Likewise, we are particularly supportive of measures in case study 16 that reference boys and the role education and supportive messaging can play to try and prevent young people engaging from harmful misogynistic content, and would push for clarity on how civil society can inform this educational messaging.

Measures set out in the good practice also feel sparse. We urge Ofcom to include more commentary around the types of harmful content that users could filter out, particularly linking to previous points around virality of misogynistic and extreme content. Content filtering should be updated constantly to respond to new trends and ways of subverting the algorithm, as users are likely to be able to circumvent filters over time.

Action 8: Enable users who experience online gender-based harms to make reports

We welcome references made by Ofcom to the way poor reporting systems can erode user trust over time, a point which is applicable to all user safety features. It is important to recognise that a lot of behaviours that make up course of conduct offences are not illegal as individual incidents, which makes it difficult to report these types of offences to platforms whose reporting is carried out on each incident. We recommend that this is noted alongside case study 23. Reporting should reflect this issue by giving users the option to report stalking or domestic abuse, which includes the opportunity to report multiple incidents at a time or allowing users to save posts as evidence and submit at a later date. They should allow for complaints about a series or pattern of communications

as well as to features of the services itself, for example, the way the recommender algorithm works, or other 'dark patterns' and nudges, or tools for creation.

The adequacy of complaints processes should be part of the risk assessment. The provider should also ensure that the design of complaints mechanisms is user-centric: that is, visible, easy to use and age and language appropriate. The regulator must also regularly assess whether such processes are fit for purpose. Regulated services must work to identify trends and developments in user reporting and incorporate this in any transparency reporting obligations to the regulator.

Reports processes should be constantly monitored in line with the feedback forms and updated to ensure they are user friendly.

We also note that the VAWG Code of Practice developed by EVAW and partners[13] is only referenced in relation to the use of trusted flaggers in case study 22. Trusted flaggers, whilst important for escalating reports of abuse when a platform has not acted swiftly, put the onus on civil society to identify and report harm, and would therefore not be recognised as the pinnacle of best practice. Where trusted flaggers are used, their roles must be evaluated and monitored for how often organisations use the feature and how effective they are for supporting victims. We refer Ofcom back to our full VAWG Code of Practice for further detail about the ambition that could be achieved in this guidance.

Action 9: Take appropriate action when online gender-based harms occur

The guidance states that "Providers often do not allocate sufficient resource and expertise to ensure appropriate action is taken when users experience or encounter online gender-based harms. Where this is the case, providers may fail to respond to user reports, and their internal content and search moderation systems may not identify harmful content." (pg. 56).

EVAW has historically called for a Tech Tax that ring fences tax collected from tech companies to fund preventative online gender-based violence work, for example 10% of the revenue raised from the Digital Services Tax ring-fenced to fund specialist VAWG sector efforts to effectively address online VAWG, with 50% of this ring-fenced for specialist 'by and for' led services for Black and minoritised women and

13 https://pdf.browsealoud.com/PDFViewer/_Desktop/viewer.aspx?file=https://pdf.browsealoud.com/StreamingProxy.ashx?url=https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/2022/05/VAWG-Code-of-Practice-16.05.22-Final.pdf&opts=www.endviolenceagainstwomen.org.uk#langidsrc=en-gb&locale=en-gb&dom=www.endviolenceagainstwomen.org.uk

girls. We reiterate the need for appropriate resourcing of internal safety mechanisms that protect women and girls on tech platforms.

In the foundational steps section, we call for a minimum standard for performance targets, including minimum standards for the amount of complaints dealt with or a review of how many complaints were dealt with effectively every 3/6 months.

There are several welcome recommendations in the good practice section, including case study 25, which introduces a strike based system to try and target repeat perpetrators. This could go further to consider perpetrators using multiple accounts, which tech platforms must become more adept at identifying.

Question 3: Do you have any comments about the effectiveness, applicability or risks of the good practice steps or associated case studies we have highlighted in Chapter 3, 4 and 5? Are there any additional examples of good practices we should consider? Please provide evidence to support your comment.

Ofcom have outlined that the foundational steps section of the guidance are those which tech companies are already legally obliged to take to be compliant with measures set out in the Illegal Harms Code and Children's Code. The guidance provides some detailed case studies relating to intimate image abuse and stalking which outline positive steps platforms can take to mitigate against this abuse, however they are included within that good practice section. Given that both stalking and intimate image abuse are listed as priority offences in the Act, and are further outlined in the Illegal Harms Code, they should instead be placed in the foundational steps, or 'minimum steps', section to reflect the duties that platform have to protect their users from those offences.

For example:

- Case study 9 - Geolocation relates to stalking so should therefore be included in the foundational steps
- Case study 12 - Intimate image abuse content should be taken down, not just deprioritised, so we urge Ofcom to go further to ensure reports of this particular content being dealt with more swiftly
- Case study 10 - Hash matching for intimate image abuse should be included in foundational steps alongside CSAM given that it is a priority offence in the OSA and outlined in the illegal harms code

Likewise, we welcomed the inclusion of Child Sexual Abuse Material (CSAM) warnings when users are searching for CSAM content that flag information and support but would

encourage this to be extended to any searches for illegal content given that platforms have a duty in the Illegal Harms Code of Practice to tackle VAWG crimes such as intimate image abuse and stalking occuring on their platform.

We also note that good practice steps are based on the current evidence of what works, which is scant. In a context where harm is often systemic within the platform, safety measures must be grounded in solid evidence. Evidence that is available may not always be scalable across all platforms, often having been tested on just one platform - long-term effectiveness should also be assessed to ensure that measures do not have unintended consequences such as retraumatising survivors, ignoring intersectional harms or relying heavily on flawed AI systems. Research on the effectiveness of safety measures must also adapt to evolving technologies, the burden of which often falls on civil society who are already under-resourced. Platforms must allocate resource for more research to be carried out into what works when combatting online harms, which is grounded in an understanding of violence against women and girls and takes an intersectional approach.

Question 4: Do you have any feedback on our approach to encouraging providers to follow this guidance, including our proposal to publishing an assessment of how providers are addressing women and girls' safety? Do you have any examples or suggestions of other ways we could encourage providers to take up the 'good practice' recommendations?

EVAW campaigned alongside colleagues in the VAWG sector for a VAWG Code of Practice that created legally binding guidelines for tech platforms to follow in order to protect women and girls using their platform. This campaign was born out of the legislation's failure to mention women and girls, and the disproportionate level of harm faced by them online. Despite widespread recognition that tech platforms need to go further when it comes to women and girls' safety, the government failed to mandate a VAWG Code of Practice in the legislation and instead put a duty on Ofcom to produce VAWG guidance.

We're concerned that, in its current status, the guidance is unenforceable due to the guidance being voluntary only. We have seen time and time again that features and products introduced by tech platforms have not been risk assessed for how they might be used to perpetrate abuse of users, particularly in relation to VAWG crimes. Too often tech providers are left scrambling to reduce harm that is already occurring on their platforms, rather than preventing it from occurring. We have also witnessed the recent rollback of user safety mechanisms on sites such as Meta, an indication that tech platforms driven by profit will not prioritise the safety of their users on a voluntary basis.

WARNING: This consultation response contains language and/or material that may be distressing

Whilst Ofcom has suggested that the status of this work as guidance allows for greater ambition in the aims and measures set out, we have not seen this level of ambition reflected in the draft guidance. There are numerous suggested measures that we welcome, particularly around external oversight, the inclusion of racially disaggregated data, acknowledging the needs of survivors through co-production, the inclusion of hashing technology and the need to consult with subject area experts. However, the guidance is based on evidence from current good practice being demonstrated platforms, which we know to be scant. Ofcom should therefore be clear that this is guidance based on current industry practice, rather than innovative new interventions. For this guidance to be truly ambitious in scope, we recommend more resource is invested in technologies that prevent harm from occurring in the first place.

Question 5: Do you have any comments on our impact assessment, rights assessment, or equality impact assessment? Please provide any information or evidence in support of your views.

Where companies choose to offer services that develop rapidly, they put upon themselves an obligation of equally adaptable risk assessment processes. Cultural attitudes, social norms and behaviours are also in constant flux, reiterating the need for companies to consider their responsibilities within a universal human rights framework. Risk management and mitigation should proceed in lock step with software and societal changes. This does not just include VAWG risk-assessments of new features, but involves continuing to risk assess for VAWG in the use or abuse of speech the use of older features.

Question 6: Do you agree that our draft Guidance is likely to have positive effects on opportunities to use Welsh and treating Welsh no less favourably than English? If you disagree, please explain why, including how you consider the draft Guidance could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.

N/A