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

ISD | Powering solutions to extremism, hate and disinformation

OFCOM'S OSA VAWG GUIDANCE CONSULTATION:

ISD CONSULTATION RESPONSE

May 2025

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

ISD welcomes Ofcom's commitment to improving online safety for women and girls and the ambition reflected in this draft guidance. However, we believe there are several critical areas where the proposed approach should be strengthened to reflect the systemic, intersectional, and rights-based nature of violence against women and girls (VAWG) online.

ISD has collaborated with the Online Safety Act Network on a shared VAWG sector briefing response on the Ofcom VAWG Guidance (submitted separately) and reiterate our support for the feedback provided. To avoid repetition, below we have highlighted certain key points that we would like to emphasise that hold particular significance for the range of harms that ISD focuses on, including terrorism, extremism, hate, gender-based violence and dis- and misinformation.

1. **Legal status and the setting of expectations for engagement by companies**

We are concerned that the guidance, as drafted and worded, could be interpreted by regulated services as purely aspirational. This risks undermining the urgency and seriousness of the harms it addresses. In particular, we are concerned that despite the serious and systemic nature of the harms identified, many steps in the guidance are framed as voluntary and "good practice". This sends a mixed signal to service providers. In practice, many of the harms outlined — including pile-ons and image-based sexual abuse — are facilitated by platform architecture, features and functionalities and require proactive design changes. Likewise, the term "foundational" does not reflect that services should *at a minimum* already be doing this.

We recommend that Ofcom:

- Clarify that service providers are expected to engage with this guidance meaningfully, and that if they choose alternative approaches, they must be able to demonstrate how their measures are equally effective in mitigating risk — consistent with broader OSA principles of proportionality, due diligence, and accountability;
- Explicitly state that many of the "foundational steps" reflect actions that services should already be taking in order to meet their safety duties under the OSA — particularly where these relate to illegal content or risks to children;
- Consider, in collaboration with DSIT, how the measures and expectations outlined in this guidance can be integrated into future revisions of statutory codes of practice, including the Illegal Content Code and Children's Safety Codes, as appropriate.

2. Recognising Human Rights impacts of silencing and discrimination

The current framing rightly identifies "content and activity" that disproportionately affects women and girls and marginalised groups, but it underplays the human rights implications of silencing and discrimination — particularly for women in public life. These harms can have chilling effects on political representation, freedom of expression, access to public life, the right to non-discrimination and participation in public affairs. A rights-based framing would help reframe VAWG online not only as a safety issue, but as a structural barrier to democratic participation, civic engagement, and equality.

Although the impacts of gendered disinformation and coordinated harassment campaigns are acknowledged under the harm area "pile-ons and online harassment" (p. 13), this important issue is not meaningfully reflected in the subsequent expectations or actions set out for service providers in the guidance. It does not sufficiently account for the unique risks faced by women and girls in public life — including for example politicians, public figures, journalists, activists and content creators — who are frequently subjected to coordinated abuse and gendered disinformation. ISD's research has shown that these harms, especially when amplified during election cycles, have far-reaching implications not only for individual safety, but also for democratic participation and systemic exclusion from digital public life. The guidance should more explicitly acknowledge these harms and provide direction to platforms on protecting users in the public sphere from systemic silencing. We provide further details and recommendations in Question 3 on this.

While we recognise that Ofcom is not responsible for the framing or reform of hate crime legislation, it is important to note a significant gap in the broader legal context in which the VAWG Guidance will operate. Currently, hate crime legislation in the UK covers offences motivated by hostility based on race, religion, disability, sexual orientation, or transgender identity. However, it does not include sex or gender as protected characteristics, meaning that misogynistic abuse and mobilisation — even when clearly targeted and hateful — may fall outside the scope of hate crime protections. This is increasingly out of step with the online landscape, where misogyny is not only prevalent but also increasingly mainstreamed, organised, and ideologically motivated — particularly among young men. ISD's research has shown the growing visibility of misogynist influencers, coordinated harassment campaigns, and radicalising ecosystems that are not fully addressed under existing legal frameworks. While changes to hate crime legislation are beyond Ofcom's remit, this gap reinforces the importance of the VAWG Guidance addressing this important harm area which currently suffers from weaker legal frameworks.

3. Strengthen the recognition of systemic platform design and algorithmic harms

While the guidance rightly identifies that online harms to women and girls are not isolated incidents, it does not go far enough in recognising platform design — particularly algorithmic recommender systems — as an active contributor to those harms. Research by ISD, including on TikTok and YouTube, demonstrates how platform algorithms can systematically guide users toward misogynistic, abusive, and radicalising content without users explicitly seeking it, contributing to a broader culture of hostility, reinforcing online and offline harms directed at women and girls. It is also important to recognise that this type of content can negatively

affect young men and boys. For example, a recent report by Movember found that engagement with so-called masculinity influencers is associated with poorer mental health outcomes among this group.

While paragraph 2.12 (page 14) acknowledges that "online misogyny often circulates amongst – and is promoted to – boys and men," this important observation is not fully developed or carried through the guidance. We recommend the guidance more clearly define algorithmic amplification as a form of harmful "content and activity" to focus on, and urge platforms not only to assess but to mitigate such risks in "foundational steps" and "good practices" — including via adjustments to recommendation systems, such as down-ranking or excluding from recommendations altogether known harmful influencers or ideologically-driven misogynistic content. These could mirror the expectations already placed on platforms under similar regimes, such as the EU Digital Services Act (Article 35), and should be implemented across all jurisdictions in which a platform operates, including the UK. In line with the OSA's risk and impact assessment duties, platforms — particularly Category 1 services — should be expected to assess how their algorithmic recommender systems contribute to online harms, including disproportionate risks as part of their illegal content and children's safety risk assessments. Ofcom should clarify that these assessments must include gendered and intersectional risk dimensions, including how algorithmic systems may amplify misogynistic influencers, ideologically driven content, or pathways to gender-based radicalisation.

4. Acknowledging intersectional risk and disproportionate impacts

Although the guidance makes reference to intersectionality (e.g. para. 2.13), this framing is inconsistently applied across the document. Many harms experienced by women and girls are shaped by intersecting identities — such as race, gender identity (including transgender and non-binary), sexuality, disability or religion — which can intensify vulnerability and limit access to redress. These dynamics are critical when assessing severity, recurrence, and platform responsibilities. This is also recognised by Ofcom's Online Experiences Tracker data, which in 2024 found that reports of stalking, cyberstalking or harassing behaviour are higher among transgender and non-binary people (16% compared to 4% of cisgender respondents). We recommend stronger integration of intersectional analysis into each category of content and activity addressed.

In addition, greater transparency is needed to monitor how harms disproportionately affect different user groups. Under the OSA, platforms are required to submit annual transparency reports when issued a notice by Ofcom. Acknowledging this echoes ISD's previous recommendations under consultation on draft transparency reporting guidance (October 2024), we recommend that Ofcom consider strengthening this framework by:

- Encouraging the inclusion of disaggregated data (e.g. by gender identity, race, disability);
- Exploring more frequent reporting on high-risk areas such as gender-based harms;
- Promoting standardised metrics to ensure consistency and comparability across services.

These measures would enable more effective scrutiny of platform responses to intersectional harms and support the development of targeted, evidence-based regulation.

5. Gendered dynamics of perpetration must be acknowledged and addressed

The guidance would also benefit from more explicit recognition that the vast majority of online VAWG is perpetrated by cisgender men and boys. While this is briefly noted on page 14, it is not meaningfully integrated into the guidance's framing or recommendations. Failing to name this reality risks obscuring the social dynamics that underpin many online harms. By acknowledging the role of male perpetration, the guidance could help inform better platform interventions, such as early detection of radicalisation pathways and risk factors in male-dominated user ecosystems (for example, in impact assessments). At present, the guidance focuses heavily on impacts on women and girls without analysing the design-level incentives that allow these harms to spread, or the user dynamics that facilitate them. This limits the effectiveness of measures seeking to "prevent harm" and reduces these to mere "harm reduction" measures.

In addition, the guidance should acknowledge that misogyny is not only propagated by cisgender men. For example, ISD's research and broader discourse show that some "trad-wife" influencers, lifestyle content creators, and wellness influencers also serve as vectors for mainstreaming harmful misogynistic narratives. These narratives may appear in more insidious, coded forms — but still contribute to the cultural normalisation of gender-based harm, especially within online spaces.

Furthermore, while the guidance is focused on violence against women and girls, its current framing does not sufficiently acknowledge that gender-based harms can also affect nonbinary, transgender, and intersex individuals, often in similar or intersecting ways. This omission may unintentionally reinforce a binary understanding of gender and of who is impacted by misogynistic abuse.

6. Small but high-risk platforms

Given the key role many smaller or medium-sized platforms play in the online ecosystem related to the perpetuation of VAWG, terrorism, extremism, hate, and disinformation, we would like to reiterate our feedback provided to Ofcom in our response to their categorisation consultation in May 2024 and on the Draft Statement of Strategic Priorities for Online Safety in January 2025:

> *"We share the concerns outlined by the Online Safety Act Network here that Ofcom is not making use of the additionally flexibility provided in the final Act that allows either size or functionality to be considered when assessing the levels of risk a service presents, and therefore the types of duties they should be subject to. Ultimately the intent of the Act is to effectively mitigate risks online, and we are concerned these proposals will leave important loopholes for certain small but high-risk services that will not be categorised.*
>
> *While recognising the requirement for Ofcom to regulate in a targeted and proportionate way, and in practice to prioritise the most risky or harmful services, we fail to understand why Ofcom would appear to restrict itself beyond the provisions in the Act. In instances*

where a service could present high levels of risk, offer relevant types of functionalities, but fall short of the user number thresholds, we would be concerned that they would escape categorisation under Category 1 or 2b under the current proposals, and therefore be exempt from important additional duties that could enhance user safety, both on and off-platform.

Ofcom correctly notes that the size of a service has a significant impact on the speed and breadth of the dissemination of user-generated content on that service. However, this does not appear to account for cross-platform dynamics and the interconnected nature of the online ecosystem of platforms and services, where harmful content or activity is often initially disseminated or coordinated on smaller platforms before migrating to larger platforms, for example in cases of targeted harassment or hate.

Without reliable public user numbers for many services, it is also very difficult to independently assess which services would be captured by the proposed thresholds, and in the forthcoming register of categorised services. This makes it difficult to assess whether the proposed approach will appropriately capture small, high-risk services that we encounter during our monitoring of platforms that play a key role in online extremism and hate."

This limited categorisation approach also has direct implications for the effectiveness of the VAWG guidance, as many of the "foundational steps" outlined in the guidance are only expected of services in certain categories. If small but high-risk platforms are not captured under Category 1 or 2B, they may not be expected to implement even baseline safety measures — despite being significant vectors of VAWG-related harms. This risks significantly diluting the intended impact of the guidance and leaving key gaps in protection for users, especially women and girls.

7. Harmful content pathways affecting women and girls must also be addressed

While boys and young men are disproportionately targeted by misogynistic influencers and conspiracy ecosystems, women and girls are also exposed to harmful online environments that shape gendered self-concepts and reinforce inequality. These harms include not only direct abuse, but also exposure to content that encourages self-harm, body dissatisfaction, disordered eating, and self-silencing — often driven by design features such as beauty filters, algorithmically promoted trends, and content loops that reward appearance-based engagement

At the same time, some women and girls are actively targeted by extremist or regressive gender ideologies, including violent pornography and content that reinforces dehumanising gender norms.

These forms of ideological socialisation — whether toward internalised harm or toward participation in gender-based hate — must both be recognised. In line with safety by design principles, the guidance should urge platforms to assess how their design features contribute to these pathways, and mitigate their effects through default settings, friction, or user empowerment tools.

8. **Misalignment with illegal content codes and risk guidance**

Finally, we note a lack of consistency across Ofcom's guidance documents. While both the Risk Assessment Guidance and the Illegal Content Codes of Practice acknowledge that certain harms — such as intimate image abuse or online harassment — disproportionately affect women and girls, these references are often fragmented and not explicitly framed as part of a broader VAWG agenda. The term "violence against women and girls" is notably absent. This lack of explicit framing risks undermining the coherence of Ofcom's approach and may reduce clarity for service providers on how to assess and mitigate gender-based harms consistently across regulatory obligations. We recommend that Ofcom move toward a more integrated regulatory approach, ensuring that expectations regarding gender-based violence are reflected across all related guidance and enforcement documents.

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

ISD welcomes Ofcom's structured approach to improving online safety for women and girls through the nine proposed actions. These actions represent a significant step forward in recognising that platform design and governance decisions shape the online experience — and can either mitigate or magnify gender-based harms. However, we offer the following observations and recommendations to strengthen the framework, including clarifications around what should constitute baseline expectations and where further specificity, or ambition, is needed.

1. **Framing and structure of the actions**

We recommend re-framing the language of the action framework to better reflect regulatory expectations:

- **"**Foundational steps" should be presented as "current expected practices" for all services with relevant features or risk profiles.
- "Best practices" should be rephrased as "future standards" that services should begin working toward. This would align the guidance more closely with Ofcom's broader risk-based regulatory model and help avoid ambiguity that could result in inaction.

Additionally, as noted in our response to Question 1, many of the foundational steps are only expected for large or multi-risk services. However, research and enforcement experience (for example, the takedown of small but high-risk sites such as Kiwifarms) clearly show that small, single-risk platforms can be significant enablers of VAWG, including image-based abuse, incel radicalisation, and stalking. These risks should be explicitly recognised, with commensurate obligations placed on relevant services.

2. **Risk assessments and recommender systems**

Action 2 encourages risk assessments "that focus on harms to women and girls," including through user engagement. While welcome, this remains vague. While Ofcom's guidance acknowledges that certain online harms disproportionately affect women and girls, there is no explicit requirement for platforms to assess the gendered impact of their algorithmic

recommendation systems. The Risk Assessment Guidance and Illegal Content Codes of Practice address VAWG and gender-specific risks in general terms but do not provide detailed directives on evaluating algorithmic trajectories and their potential to amplify misogynistic content.

The guidance should:

- Require platforms to assess how their recommender systems contribute to VAWG, including gendered exposure to hate, harassment, and radicalising ideologies, and in this include an intersectional lens to consider the profiling and disproportionate impact on specific groups based on race, gender identity, disability and or religion.
- Ofcom's Children's Risk Assessment Guidance mandates that platforms assess risks related to content harmful to children, including content promoting unrealistic body images, which can lead to body dissatisfaction, eating disorders, and mental health issues such as suicidal ideation. Platforms should ensure that their risk assessments for child safety explicitly consider these factors, evaluating how algorithmic systems may contribute to such harms.

While Ofcom's current guidance emphasizes the importance of risk assessments, there is a need for clearer directives requiring platforms to adapt their algorithmic systems when risks are identified. Platforms should be encouraged and incentivised to implement changes to mitigate identified risks and to report on these adaptations, ensuring transparency and accountability in their efforts to protect users from harm.

3. Action 5: Safer defaults

We support the use of proactive safety-by-default settings, particularly in moments of acute risk, such as during coordinated harassment or when a user is experiencing a pile-on. These features — for example, limiting who can reply or temporarily restricting visibility — can be effective short-term tools for regaining control and safety.

However, it is important to acknowledge that some default settings — such as switching accounts to "private" or significantly restricting visibility — may create longer-term trade-offs. For many women, particularly those in public life (e.g. politicians, journalists, influencers and activists), such measures may not be viable without undermining their professional presence or access to audiences. Over time, these settings can contribute to a chilling effect, forcing women to self-censor or withdraw from online spaces altogether in order to avoid abuse.

Recommendation: The guidance should:

- Encourage platforms to design safety features that enhance agency and optionality, rather than assuming visibility must always be restricted.
- Support adaptive tools that allow users to dynamically adjust their level of exposure without sacrificing participation or reach.

- Include feedback from women in public-facing roles during feature design and testing, and reporting mechanisms, to ensure that safer defaults do not inadvertently limit civic engagement.

4. Image-Based Sexual Abuse (IBSA) and AI/Apps

The draft guidance rightly identifies image-based sexual abuse (IBSA), including intimate image abuse and cyberflashing, as a key harm. However, it does not sufficiently address emerging AI-enabled threats such as "nudification" or deepfake non-consensual intimate imagery generation apps, nor the broader "supply side" ecosystem through which such tools are accessed and monetised — including search engines, hosting services, and potentially app stores.

These AI tools, many of which can be used to create non-consensual intimate images (NCII), are often not categorised as user-to-user or search services, and are therefore likely to fall outside the immediate scope of Ofcom's regulation. Nonetheless, they pose a fast-evolving risk to women and girls, especially when discovered and accessed via search engines or third-party app stores.

Although app stores are currently not directly in scope under the OSA, the Act provides the Secretary of State with delegated powers to bring them into scope following Ofcom's forthcoming review (due between January 2026 and January 2027). The final guidance should therefore:

- Acknowledge the role of app stores in the NCII supply chain, including their potential use to distribute nudification or image-generation apps.
- Signal to platforms that these vectors of harm must be monitored, and should be subject to safety duties if significant risk is confirmed.

Similarly, search engine providers must ensure they are not facilitating NCII abuse, for example by indexing or promoting websites and tools that enable image-based sexual abuse. While some search engines are making progress in downranking or delisting harmful sites, the guidance should go further in clarifying:

- That search-based discovery of nudification tools constitutes a form of facilitation, and
- That risk assessments and content moderation measures should explicitly include these forms of AI-enabled IBSA.

We also encourage Ofcom to make use of its existing powers to disrupt access to and monetisation of harmful services. For example, if a provider is found to be non-compliant with OSA duties (including age assurance or NCII removal), Ofcom may apply to the courts for an order requiring payment providers, advertisers, or ancillary services (such as PayPal) to cease working with them — a critical tool for undermining the business model of persistent abusers.

We also note the ongoing legislative developments regarding deepfake or generative AI sexual content (e.g. the forthcoming Crime and Policing Bill that aims to make it a criminal

offence to create sexually explicit deepfakes without consent). The final Guidance should reflect these developments.

5. **Data, evidence and future inclusion in statutory Codes**

Some of the actions (e.g. abuse tracking, risk assessments, interventions) may be difficult for smaller providers without better data infrastructure. Rather than exempting them, the guidance should:

- Encourage proactive data collection and evidence-building, and
- Make clear that good practices in this guidance may become part of future statutory codes.

This message would help create regulatory certainty and incentivise forward planning, particularly for platforms subject to similar obligations in other jurisdictions.

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 these nine action areas? Are there any additional recommendations of good practice we should consider, or any service providers who are currently implementing similar practices that we have not included? Please provide evidence to support your comment. *Note: we are not consulting on the foundational steps and associated case studies as the measures and guidance which underpin the foundational steps have gone through separate consultation processes.*

ISD welcomes Ofcom's efforts to outline practical good practices and illustrative case studies to support service providers in mitigating violence against women and girls online. We appreciate the emphasis placed on platform design and structural prevention measures, rather than relying solely on reactive content moderation. However, we identify several gaps, risks, and areas where good practice could be strengthened or expanded to reflect emerging challenges and evidence.

1. **Election periods and crisis events must be addressed as high-risk contexts**

While the guidance identifies pile-ons and coordinated harassment as one of the key harms to focus on, it does not provide sufficient recognition of when and how such attacks escalate, particularly in moments of heightened public visibility. ISD's research has consistently shown that election periods, moments of national crisis, and high-profile legal events are flashpoints for coordinated abuse, disinformation, and VAWG against women in public life.

Recommendation: Under Action 6 (*Reduce the circulation of online gender-based harm*, para. 4.30), and Action 7 (*Support women and girls to stay online safely*), platforms should be encouraged to:

- Develop event-specific crisis protocols,
- Monitor and escalate response efforts during key civic moments, and
- Establish cross-platform and law enforcement and relevant authorities coordination mechanisms to address coordinated gendered abuse (elaborated further in points 2, 3 and especially 4 below).

2. User reporting systems must be usable and integrated with law enforcement needs

In Action 8, the guidance encourages the development of accessible reporting mechanisms for users experiencing online gender-based harm. This is a positive step, but we note two key gaps:

- First, many user reporting tools are poorly designed for VAWG-related harms. They often lack options for reporting nuanced or multi-faceted abuse (e.g. coordinated harassment, image-based abuse, impersonation), offer no facility for bulk reporting, or don't allow users to report the context of the abuse (e.g. in the midst of a pile-on). Without these features, reporting can feel futile or re-traumatising.
- Second, law enforcement engagement is underdeveloped. The guidance should promote standards that ensure platform reporting outputs are compatible with law enforcement workflows and recognise the online–offline continuum of harm — particularly for domestic abuse, stalking, or threats.

Recommendation: Expand Action 8 to include:

- User-centred design of reporting systems (co-created with survivors),
- Bulk-reporting capabilities and context-aware options,
- Structured outputs that can be shared with relevant authorities (e.g. through formal escalation protocols).

3. Cross-platform coordination and supply-chain responsibility are essential

Gendered abuse does not remain confined to one service. Particularly for intimate image abuse, extremist content, or harassment campaigns, perpetrators often use multiple services in concert (e.g. niche image-hosting forums, mainstream social media for amplification, encrypted messaging for coordination).

Recommendation: Embed in Action 6 and Action 9 stronger expectations that platforms:

- Engage in cross-platform coordination,
- Share threat intelligence with trusted parties, and
- Address "supply chain" harms, including apps or tools (e.g. de-nudification apps) that facilitate VAWG but may fall outside direct platform control (see our point under Question 2 for detail).

This aligns with broader discussions in digital policy around supply-side accountability and builds on lessons from the similar and established cross-platform coordination initiatives: such as through the EU's Code of Practice on Disinformation (its "Task Force", and Measure 16.1: "Relevant Signatories will share relevant information about cross-platform information manipulation, foreign interference in information space and incidents that emerge on their respective services for instance via a dedicated sub-group of the permanent Task-Force or via existing for exchanging such information") the Christchurch Call to Action, or the Global Internet Forum to Counter Terrorism (GIFCT).

4. Additional good practice recommendations

We encourage Ofcom to consider the following additions to the current list of good practices and service examples:

- Transparency dashboards for gender-based harm: Platforms should publish regular updates on metrics such as time to removal, prevalence of abuse against public figures, and algorithmic exposure to misogynistic content — disaggregated by gender where possible. The monitoring, measurement, and transparent reporting of VAWG by platforms are prerequisites to understand and explain the nature, scale, and scope of the phenomenon. Additionally, API access to publicly accessible data should support public interest research and enable evidence-based decision-making. The Global Partnership for Action on Gender-Based Online Harassment and Abuse ("the Global Partnership") together with UN Women, the WHO, UNFPA and UNICEF initiated efforts toward enabling the production of accurate, reliable and comparable data and knowledge around Technology Facilitated Gender-Based Violence (TFGBV). The UK is a driving member of the Global Partnership, and should ensure these international efforts are reinforced and implemented at home.
- Audit logs for algorithmic changes: Platforms should maintain records of algorithmic interventions (e.g. downranking influencers or keywords), and share these as part of their risk mitigation reporting.

Question 4: Do you have any feedback on our approach to encouraging providers to follow the 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?

ISD supports Ofcom's commitment to promoting uptake of the guidance through transparency, public accountability, and recognition. Given the guidance is non-binding, the success of implementation will depend significantly on how well Ofcom can shape incentives — both reputational and practical — to ensure meaningful engagement by regulated services. We welcome the proposal to publish assessments of how providers are addressing the safety of women and girls and offer the following recommendations to strengthen Ofcom's approach.

1. Develop a dual-layered public assessment mechanism

We support Ofcom's proposal to publish an assessment of service provider compliance with the VAWG guidance and recommend this take the form of a dual-track reporting model, grounded in the regulatory duties of the.

- A **public-facing "report card"** with clear indicators of platform safety and good practice (e.g. safety-by-design features, responsiveness to abuse reports, support for women in public life), tailored for journalists, advertisers, civil society, and end users. These indicators should build on the risk factors set out in Ofcom's *Risk Assessment Guidance and Risk Profiles*, and reflect services' duty to consider the impact of harms that disproportionately affect women and girls. Categories could include:

  - Safety by design (e.g. friction measures or protective defaults),
  - Responsiveness (e.g. average review time for VAWG-related reports),
  - Trust and safety resourcing (e.g. number of moderators per user or language),
  - Algorithmic risk and impact assessments, including gendered recommender system audits and mitigation measures.

- A **detailed technical transparency annex** providing platform-specific data on concrete, comparable performance indicators — enabling scrutiny by regulators, researchers, and civil society. These metrics should be linked to known risk factors, intersectional harm profiles, and structural drivers of abuse (e.g. amplification of misogynistic influencers, coordinated harassment). Further, as detailed in our answer 1.4 above, greater transparency is needed to monitor how harms disproportionately affect different user groups. Specifically, this should include: the inclusion of disaggregated data (e.g. by gender identity, race, disability) and standardised metrics to ensure consistency and comparability across services.

Such an approach would help align the UK's safety framework with the EU Digital Services Act's Art. 42 dashboard and systemic risk assessment model, while also enabling public accountability and cross-sector pressure on non-compliant platforms. It would also support compliance with equality principles under the Equality Act 2010, ensuring that disproportionate harms based on sex, gender reassignment, race, or disability are not only acknowledged but monitored and mitigated through transparent reporting.

2. Incentivise adoption through external validation and industry benchmarking

Based on ISD's experience working with platforms, public benchmarking — especially when tied to external research, cross-sector initiatives, or advertiser advocacy — can be a powerful motivator. For example:

- Platforms are more likely to prioritise safety improvements when civil society and media scrutiny is facilitated by high-visibility metrics.
- External rankings (e.g. *Ranking Digital Rights*) have historically driven policy improvements.

Ofcom's own assessments could be reinforced by allowing trusted external auditors to conduct independent evaluations of platform adherence to the guidance. To ensure that external audits are meaningful rather than symbolic, Ofcom should:

- Clarify who may serve as a trusted auditor, including civil society organisations with demonstrated expertise in online safety and gender-based harms, academic institutions with ethics oversight, and select independent research bodies with safeguards in place for user privacy and data integrity.

3. **Promote taxonomy-based standard setting and cross-platform alignment**

Linking to Ofcom's Risk Profiles as per its "Risk Assessment Guidance and Risk Profiles", we recommend that Ofcom develop standardised taxonomies of gender-based online harms — distinguishing clearly between:

- Illegal content (as already indicated in this Guidance, though with further breakdown),
- Legal-but-harmful activity (e.g. misogynistic influencers, gendered disinformation), and
- Structural or systemic harms (e.g. algorithmic amplification, design patterns).
- Context-dependent/borderline harms (e.g. satire, coded language, dog whistles), which often evade moderation but still contribute to gendered toxicity)

A clear taxonomy would:

- Help platforms operationalise the guidance through safety engineering and moderation teams,
- Enable better cross-platform coordination, while recognising that platforms differ in scale, resources, and policy enforcement capacity, and
- Support alignment with international efforts, such as those led by ISD and UNFPA on VAWG harm classification and safety protocols.

Such taxonomies could also be used in transparency reporting, algorithmic audits, and independent evaluations.

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.

ISD welcomes Ofcom's inclusion of an impact assessment, rights assessment, and equality impact assessment as part of the draft guidance. However, we believe these assessments would benefit from greater specificity, mainstreamed intersectional framing, and more robust engagement with the structural drivers of online VAWG, particularly regarding platform design and algorithmic systems.

1. **Freedom of expression and interdependent rights require deeper treatment**

While the rights assessment notes the need to balance Article 10 (freedom of expression) of the UK's Human Rights Act (1998) with the right to safety, privacy, and participation, this discussion remains limited and abstract. We recommend a clearer articulation of:

- The specific groups whose expression is most at risk due to online abuse, such as women, girls, journalists, politicians, and activists, and marginalised communities e.g. LGBTQ+ and queer communities. The human rights costs of inaction — particularly for women and girls who are driven offline by abuse, including journalists, politicians, and activists.

Failure to act on systemic VAWG can amount to a de facto suppression of expression and participation, particularly for already marginalised groups. The rights assessment should reflect this, using an affirmative rights lens to clarify that action is not only compatible with human rights — it is required by them.

2. Equality impact assessment needs strengthening on gender and intersectionality

Although the guidance refers to women and girls as disproportionately affected (and cites protected characteristics under the Equality Act), the equality impact assessment does not mainstream intersectionality throughout.

ISD recommends the assessment be strengthened to:

- Acknowledge how VAWG is shaped by intersecting identities — such as gender identity, race, disability, sexuality, religion, and age — which influence both exposure to harm and barriers to redress. Intersectionality should be reflected not just in user experiences but also in how Ofcom evaluates platform practices (e.g. moderation gaps, language bias, reporting failures for marginalised groups)
- Ensure this lens is integrated into the risk profiles, good practice steps, and platform reporting expectations, rather than mentioned only at a high level.

3. Algorithmic systems are under-assessed across all three evaluations

Given Ofcom's own recognition of the role of algorithmic recommender systems in amplifying VAWG-related content (pages 27, 31 and 33), it is notable that these systems are not mentioned in the rights or equality assessments, nor are the implications of not intervening in their design.

To ensure the impact assessments are complete, we recommend:

- Including explicit analysis of the freedom of expression implications of algorithmic systems that prioritise and amplify misogynistic content, even if the content itself is technically legal.
- Referencing alternative measures to removal — such as algorithmic deprioritisation, demotion, or insertion of counter speech — as less restrictive means of achieving a rights-protective outcome.
- Drawing on international comparators — such as the EU Digital Services Act (DSA) Articles 34 and 35 — which require systemic risk assessments of algorithmic recommender systems and mitigation planning, particularly for gendered harms.

This would clarify that recommender system audits are a proportionate and rights-compatible intervention, and that failure to address their role in shaping harm may itself have equality and rights implications.

4. Monitoring and transparency obligations should be embedded

To support the implementation of the guidance and allow for meaningful assessment of its impact, the equality and rights assessments should include commitments to:

- Track disaggregated data on implementation and platform responses (e.g. by gender, race, role in public life),
- Regularly review the guidance's real-world impact, for example on a five-year basis, and
- Integrate these insights into Ofcom's broader transparency and enforcement framework under the Online Safety Act.

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.

While ISD has not undertaken direct research on Welsh language harms, our international work highlights the risks minority-language communities face when platforms under-resource moderation and safety processes. This can lead to uneven enforcement of safety standards, underreporting, or delayed responses to harmful content in these languages.

To maximise the positive effects of the guidance and ensure equality of treatment for Welsh users, Ofcom could:

- Encourage platforms to report on language-specific content moderation resources — drawing from the EU Digital Services Act (DSA) model, which requires platforms to disclose the number of moderators per EU language (Article 42).
- Consider broader applicability to other underrepresented UK languages (e.g. Irish, Scottish Gaelic), particularly where harmful content circulates across devolved regions.

Transparency in linguistic resourcing would help ensure that Welsh users — especially women and girls — receive the same level of safety and protection as users engaging in English-language environments.

While the current question focuses on Welsh, we also note that many widely spoken non-English languages in the UK — such as Polish, Urdu, Punjabi, Bengali, and Arabic — are likely under-resourced by platforms in terms of content moderation and trust & safety operations. As with Welsh, this can result in uneven enforcement of platform rules, reduced protection from harm, and limited access to redress for large user groups. Ofcom should consider encouraging platforms to publish disaggregated moderation data for other frequently used UK languages beyond English.