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

Consultation Response Overview

Where the Ofcom Guidance sets out a range of methods for tech platforms to deal with Violence Against Women and Girls (VAWG)* in internet spaces, *prevention* should remain the core focus. Effective tackling of online abuse and violence involves effective detection and removal in the first instance, and then a range of meaningful safeguards for users where VAWG has already occurred. To do this, comprehensive approaches are needed to maximally identify and capture abusive content as heterogeneous - we recommend multi-query assessments that practice intersectionality, comprehensive typologies, and consider a range of actors and variables (that determine and define the different contextual manifestations of OGBV), with special interest in emerging and neglected forms of OGBV. Additional to this, responses to OGBV that employ user-driven design, which reflect users' needs, and that fundamentally centre harm-reduction as the 'golden rule' running through all interventions.

*Note that, we, *Equally Safe Online* (ESO) refer to Online Gender-Based Violence (OGBV) rather than just VAWG.

Who We Are

Equally Safe Online (ESO) are submitting to the *Ofcom Consultation on draft Guidance: A safer life online for women and girls* (referred to as 'the Guidance' in this document) following on from our attendance at the Ofcom Scotland consultation event in May 2025 for tech and VAWG practitioners.

Aligned with the Scottish Government's *Equally Safe* strategy, ESO an interdisciplinary, collaborative project where social science meets Machine Learning and Natural Language Processing (NLP). Experts in safeguarding, gender-based violence (GBV) and digital education from the University of Edinburgh School of Informatics, School of Social and Political Science (SPS), and the School of Education are working with computer scientists at Heriot-Watt University (HWU) to build a user bot for prevention, intervention, and support in online gender-based violence (OGBV), with the project scope specifically being discourse-based OGBV on social media platforms. However, the ESO project has increasingly studied the inseparability of image-based and language-based content in OGBV. We encourage Ofcom's Guidance to consider the complex ways that violence and hate crimes manifest through this combination, such as abusive memes, commentary on/narrative attached to images, and the use of messages to carry abusive images.

ESO's research methods entail intersectional participatory grassroots design with third-sector partners and victim/survivor stakeholders. Co-creation focus groups, both online and in-person, were conducted with local and national feminist charities, including Rape Crisis, Women's Aid, SafeLives, EmilyTest, and Glitch (the UK's only OGBV charity) as well as Children's Rights charities such as YoungScot. Sessions with young people (YP) included general population mixed gender groups and YP not in education, supported by Rape Crisis specialist YP workers, teachers, and youth workers.

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

To create a programming framework, ESO's first work package (WP1) created a taxonomy – or classification system – for OGBV by posing the question 'what does OGBV look like?' to participants, and by considering existing law and policy, including the Online Safety Act 2023 and the Equality Act 2010. The latest iteration of the taxonomy is linked here. The taxonomy, on an ongoing basis, is tested against thousands of real-world social media examples. Used as the framework of an annotation scheme to label OGBV instances and scrutinise annotators' interpretations and perceptions, it is continuously assessed and innovated. The ESO project comes to an end in November 2025.

A taxonomical approach allows for artificial intelligence (AI) solutions to detect GBV, to empower internet users to filter instances of hate and abuse, as well as to permit the production of 'counter-speech' (CS), generating responses such as fact-checking, signposting, and bystander help, in line with the appropriate detection. This is because the type of GBV detected needs to be followed by a pipeline of interventions in accordance with the specific type detected - all of this requires robust, bespoke assessment of what is specifically manifesting in an instance of OGBV. In focus groups, we brought AI-generated examples of CS to test the different parameters and possibilities of CS and ask participants what best practice would look like dependant on the contextual incident(s) of OGBV.

Our taxonomy is a first for the computational field, as an OGBV taxonomy does not exist; existing taxonomies focus only on the concepts/labels of sexism and/or misogyny.

Considering high levels of youth victimisation, our second work package (WP2) invited young people (YP) to participate in the development of the taxonomy and advise on the detection of OGBV, CS, and OGBV typologies. Using creative materials, this data-collection prioritised young GBV survivors and including a mix of gender, geographical location, socioeconomic background and a group not in mainstream education.

Taking WP1 and WP2 together, ESO's focus groups took place over 1.5 years, facilitating challenging discourse across locations and audiences in Scotland and the wider UK, from children to experienced practitioners coming from a variety of occupational roles, mainly frontline service workers and policy. Fieldwork discussions included gathering testimonies on what works to effectively respond to OGBV and providing insights from those particularly at risk and with direct lived experience.

The ESO project is now at the dissemination stage, building our user bot, which utilises our taxonomy, writing publications on areas like methods and ethics, and applying for art community outreach funding. We are undertaking a range of exercises to access datasets, train our models with real-world examples, and user-test with annotators.

We recently won the Best Impact from a Data-Led Project award from the University of Edinburgh Centre for Data, Culture, and Society (CDCS) Digital Research Prizes 2025 and have been granted an Impact Acceleration Award (IAA) to build a 'Support Buddy' application for OGBV. ESO has presented to Ministers at the Scottish Government Policy Forum, and we are in talks with national partners to produce educational materials in line

with the Equally Safe at School (ESAS) programme, the *Equally Safe* strategic arm for secondary education. The project has received letters of support from Ofcom, Members of Scottish Parliament (MSPs) and the Scottish Government's VAWG team, the Equality, Inclusion and Human Rights Directorate and third sector organisations such as Not Your Porn and the Revenge Porn Helpline.

ESO's Research Contributions

Sprint 1

Work package 1 (WP1) of ESO was dedicated to constructing the frameworks to inform AI modelling. ESO's bot will be programmed using our taxonomy categorisation system, applying a multi-query framework to detect different OGBV.

Where ESO utilised co-creation and participatory methodologies to produce the taxonomy, we strongly support Ofcom's frequent emphasis on co-production throughout the Guidance. It is particularly encouraging that Ofcom mandates 'abusability' testing and safer default settings. By embedding protections from inception, and having users guide and tangibly benefit from these protections, Ofcom positions tech firms to move beyond ad-hoc moderation and have more systematic approaches grounded in users' real concerns. However, Ofcom's user-centric approach needs to retain equal and/or heavier focus on firms possessing ultimate leverage to stop and redact harmful content (as arguably the most pertinent user concern over their ability to control content). In other words, having a 'harm is harm' approach, regardless of what users can do once violent content is circulated. OGBV is often regulated in ways overly reliant on user self-protection. Yet, there are other areas where user involvement can be utilised for tackling OGBV. Where ESO's research with internet users focused more on their lived experiences, we were able to study online subjectivities and contexts. From this, we recommend that user interventions should be built with contextualisation in mind. For example, where healthy discourse on OGBV should not be caught up with OGBV itself, where cryptic language (code/flag words, euphemisms, hints) indicate OGBV rather than overt forms/open display of OGBV, and where OGBV is rarely isolated and one-off, and sits across behavioural patterns and trajectories that need identifying.

In addition to the taxonomy, a typology of OGBV was produced in conjunction, inclusive of less-visible, less-considered, and emergent OGBV forms such as body shaming, sexual grooming, and rape myths. ESO's focus group with Amina (the Muslim Women's Resource Centre) discussed how honour-based violence (HBV) on social media platforms – for example, a male ex-partner posting a picture of a Muslim woman without her head covering - is often missed by rules and moderators as GBV due to poorly diversified typologies. Other focus groups discussed how online trends occur in cycles of resurgence and repackaging. This can be tracked through malleable typologies – for instance, age-old gender stereotypes have been 'rebranded' through contemporary online reconfigurations, such as 'lad' cultures,

'trad wives', 'alpha males', and 'high-value men/women'. Diverse typologies, therefore, can be used as a tool to gauge ever-shifting online movements.

As ESO works in the Scottish context, we adopt the Scottish Government's *Equally Safe* strategy, which uses the multifaceted umbrella concept of GBV over siloed focuses on, for example, sexual violence. OGBV (over, for example, online sexism/misogyny) is a nascent focus in academic literature but arguably provides the most modern approach to the study of online violence and abuse. We welcome how the Guidance refers to 'gendered online harms' in addition to VAWG. We also welcome how the Guidance throughout states that OGBV, and responses to OGBV, should both be manifold. User empowerment online, and freedom from OGBV, means multiple things to users. Thus, firms should offer numerous user options, mirroring how OGBV is multitudinous in users' lives both in terms of types and occurrence levels. The multiplex construction of both OGBV itself, and OGBV solutions, should be expanded and deepened in the Guidance – for instance, where it mentions reporting, it should recognise the realities that OGBV is constant for some users.

ESO's work practices a continuum-based understanding of OGBV in order to be exhaustive of all OGBV types, configuring OGBV as pluralistic, multiplicitous, and cross-domain, and OGBV risk as lifelong and always oscillating. We encourage the Guidance to use an OGBV spectrum in order to capture, reflect, and make important categorical distinctions, between 'everyday sexisms' through to acute forms - this is needed for the likes of practical risk assessments but also to tailor specialist support and produce best outcomes for victims/survivors. Here, the Guidance also needs to be more targeted around its recommendations towards acute OGBV forms, such as sex trafficking and grooming. Furthermore, where the Guidance addresses 'harms not considered solely VAWG', but which are gendered, more recommendations are needed for these distinctly harmful and complex forms of violence. For instance, modern slavery, gang recruitment, and forced prostitution. The Guidance needs to better underline the role that firms play in upholding the most acute OGBV, as well as the various laws that come into play (such as the multiple hate crimes presented in intersectional forms of abuse) on top of the Online Safety Act.

Linked to this, the Guidance needs to address how firms play an important role in the de-escalation of the most serious outcomes offline (such as assault, suicide, and homicide), in evidence-collection around these outcomes, and in cooperation with authorities. Where Ofcom delineates between lower and higher-risk spaces, the Guidance needs to address the permeation of acute OGBV forms in sites that may be considered as lower-risk, as well as how OGBV often combines higher and lower-risk types and spaces simultaneously and can move between these. Furthermore, it needs to name how higher-risk sites are often embedded in lower-risk (for example, through links in comments) and how everyday conversations between users can quickly be multi-spatial and move swiftly up and down scales of risk as a result of lower-risk spaces being a host or gateway for higher.

Furthermore, the Guidance needs to better map typologies onto cohorts. For instance, this includes considering evidence around the forms of violence YP are disproportionately

subjected to. Whilst the Guidance recurrently mentions marginalised groups and Protected Characteristics, as well as categorical approaches towards OGBV, it needs to link these explicitly. For instance, more overtly naming the associations between OGBV forms and protected groups, such as ethnic/religious minorities and honour-based violence (HBV), and working-class communities with sexual exploitation. Applying a taxonomical approach, we encourage the Guidance to take a multi-query approach to shift from a more 'generic' assessment of VAWG (and subsequently, generic intervention and support suggestions) to articulating more multi-pronged flowcharts of 'categorisation-to-responses' pipelines, according to how the OGBV is labelled.

Moreover, the Guidance does not map typologies, cohorts, and platform types. OGBV manifests differently across, for instance, gaming sites, dating sites, conversation feeds, and platforms that are primarily image-based (e.g. Instagram) rather than discourse-based (e.g. Reddit) platforms. These variant online ecosystems are made up of different demographics (for instance, certain occupations use X and LinkedIn, and different generations are more likely to use Facebook over Snapchat) and this shapes the different risks, and thus the different intervention types and what efficacy looks like.

ESO divided the structure of stakeholder focus groups into firstly, taxonomical design and secondly, creating counter-speech (CS) strategies. In regard to these two areas, the Guidance should consider the following.

- Concerning counter-speech, the Guidance should pinpoint the different CS types firms could consider in their platform functionalities and encourage them to map out and enhance where CS is/could be available - firms know their own unique platforms best. Good practice examples of CS (for instance, signposting to user functions and/or to provisions internal and external to the platform) need to be more explicitly made available in the Guidance.
- ESO considered CS that addressed the three groups of victim/target, bystander, and perpetrator. The Guidance needs to more comprehensively address these three groups and have a focus on (the links between) mental health and GBV. This includes signposting support for victims and information for perpetrators, such as those who wish to escape harmful groups and those concerned about their behaviours or the content they are exposed to.
- In various places across the Guidance, reference is made to future human evaluation of how the Guidance has landed - however, how this research would cover these aforementioned multiple groups is not entirely clear, and thus it needs to systematically consider different spaces and people. Precise metrics and good practice examples (for the likes of user surveys) are not entirely unpacked. This includes what success benchmarks look like, and importantly, how these benchmarks intersect with regulator/Government targets. The following are not mentioned in the Guidance: the regularity of such evaluative research, its sustainability, accountability to undertake it; longitudinal monitoring; and how such insights will inform (re)modelling of policy and platform functionality, and be utilised by academics and Government

- The Guidance needs to strongly consider the role that bystanders play in achieving the Guidance's aims. In addition to facilitating community empowerment, resistance, and peer support, the Guidance needs to consider the body of evidence on perpetrator change. Where the Guidance perhaps makes more broad-sweeping suggestions about firms engaging in data-sharing, this is a clear opportunity for the Guidance to make more specific recommendations on the data that academia, policy, and practitioners need - namely, that on what interventions (and CS) lead to hard positive outcomes, such as take-up of support and reductions in perpetration. Firms possess missing data that would help fill urgent knowledge gaps - namely, there is a dearth of data around 'what works' on effective OGBV behavioural change trajectories.

Sprint 2

As stated, ESO's WP2 entailed focus group workshops with young participants. As well as many of the points above in relation to detection and mitigation, we would like to highlight some key findings relevant to the youth OGBV context. YP use specific words, phrases, and emojis to identify and discuss OGBV. This evolves frequently as the language YP use changes, as do the steps that are necessary to evade detection algorithms. The Guidance needs to take account of this and ensure there is ongoing involvement of YP in keeping abreast of changing landscapes. YP identified the importance of the context of the post, the conversation and who the poster is. They also identified the importance of appealing to bystanders in counter-speech, in ways that resonate and connect with them (e.g. facts and 'sass') and the importance of considering, and responding differently to, YP who were in danger of radicalisation, as opposed to established, often adult, posters of misogyny. They stressed that support for victims, mainly girls and women, should focus on solidarity as well as support. We encourage the Guidance to push demographics-based understandings of who predominantly uses their platforms, mapping out risks, needs, and responses accordingly. Methodological and ethical frameworks for conducting user research with YP and other vulnerable groups are not mentioned in the Guidance. Online youth-focused information, such as YoungScot 'That's Not OK!' and NSPCC information, need to be updated in light of current understandings and linked to the guidance. YP in ESO's workshops expressed the need for creative informal non-didactic educational spaces to help them reflect on their own online activity and ways of responding to online GBV, and this is an area we are keen to co-develop with YP and partners.

Sprint 3

Our Support Buddy software 'the bot' will initially be a browser plugin to detect and mitigate OGBV in real time found on popular social media websites. Fitting with Ofcom's user-protective approach, we encourage the Guidance to facilitate and promote a more formalised suite of tools produced by academia and the third sector, to bring together the rich body of (existing) self-protective measures, increase access for those in need, and bolster women and girls' informed choices. However, the serious limitations of user-protective approaches - namely, that powerful algorithms drive OGBV virality, that user protections place heavy

burdens on victims, and that interventions tend to be retrospective 'when the damage is done' - are not covered in the Guidance.

General Guidance Feedback

Specific Comments

Where the Guidance states point 6. under *Preventing Harm* (page 4), reducing circulation of online gender-based harm should be made no. 1 in this list above safer defaults. User-centric, safety-by-design should universally prioritise reduction over reactive (post-event) responses.

On the same page, under *Taking Responsibility*, the Guidance needs to explicitly name and refine data-driven approaches. The keywords of this section – governance, accountability, risk, engaging, user surveys, transparency, information-sharing, effectiveness – are all underscored by firms grasping their own OGBV data and instrumentalising that data into informing their aims, priorities, strategy, research, provisions, and operations. The direct link between data and engineering, as a solution to OGBV, needs to be stated.

In 2.8, the Guidance focuses on four key areas of harm: online misogyny, online harassment and pile-ons, domestic abuse, and image-based abuse. This covers a wide spectrum, capturing both illegal and harmful-but-legal behaviours. The inclusion of examples and sub-types, like AI-generated deepfakes, cyberflashing, coercive control, and stalking shows awareness of the breadth of OGBV and of modern threats. However, to an extent, this still misses an opportunity for a more holistic coverage of gendered harms, and it is not clear why the Guidance stops at these four categories. Whilst these four groupings offer some extensiveness, it is not exhaustively intersectional. How an intersectional approach will be enacted is elaborated on in the likes of 2.13, 2.49, and 2.55 (for instance, through disaggregated data) but the Guidance does not state how intersectional approaches marry up with "assess[ing] the risk of illegal harms" (page 3). Specifically, it does not advise, or provide, a specific risk assessment framework, and how an intersectional assessment (for example, of racial and gendered abuse combined) reshapes and enhances the illegality of harms (i.e. hate crimes always inherently involve the violation of multiple policies and laws at any one time), within a wider framing in online hate speech over just online VAWG. Linking to this, the Guidance lacks information on how firms' responsibilities map tangibly onto criminal justice processes.

The Online Safety Act also has a hierarchical approach - namely that Category 1 sites and content have more duties (of care) applied to them, as the highest-risk domains of harm. The Guidance should state how an intersectional breakdown of GBV maps onto these categories - for instance, how higher-risk harm/spaces relate to compounding crimes affecting the most marginalised users, such as combined race and gender discrimination. Here, the multidimensionality of OGBV (illuminated by an intersectional lens), should be crystalised in the Guidance's OGBV categories and in the Guidance's reference to the Bill categories.

The Foreword, providing the overall framing of the Guidance, correctly states that firms have a "duty to protect all users from this material, taking down illegal material once they become aware of it" (page 3) in line with the Online Safety Act. We encourage the Guidance to do more to state at the beginning that the law is binding but that best practice measures are a part of this mandate - indeed, the Online Safety Act dictates that companies proactively assess and manage risks ahead of the curve, going beyond merely a reactive role (such as reacting to reports) and shifting to proactive. The Guidance needs to address how intersectionality will be assessed in these terms, setting out how risk and proactive plans/measures sit across different loci of marginalisation, social groups, and GBV types.

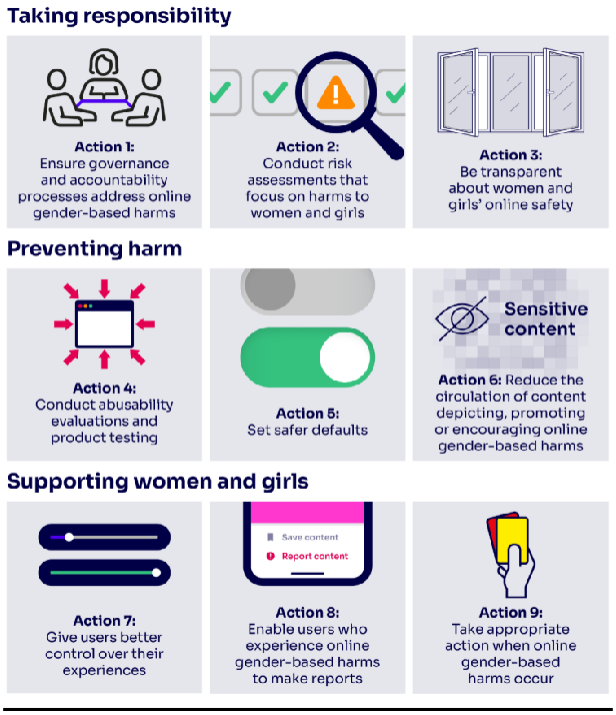
Regarding "pile-on culture" (page 3), more detail is needed on how firms and regulators will monitor feeds, threads, and patterns of activity as well as individual posts. The use of AI models to automatically detect OGBV can play a significant role in automating the process by redacting, or "flagging" content, and also provide more nuanced information as to why particular inputs are harmful. This can be used to directly inform moderation teams, protect victims/survivors, and inform bystander users.

Page 3 additionally mentions that "misogynistic speech is often not illegal, but, at scale, it can normalise harmful beliefs in boys and men". Radicalisation, and relevant legislation and governmental policy, are not mentioned in the Guidance.

Where on page 4, the Guidance states "taking greater responsibility *at all levels* for women and girls online safety", this needs to account for signposting, asking firms to deliver support offers (for instance, embedded in their reporting systems based on frequent patterns, e.g., the language used, or the form of GBV, of automatically detected OGBV in threads and posts) that take victims/survivors to practical online and offline help in their everyday lives. Where the Guidance promotes a safety-by-design approach "demonstrating how providers can embed…throughout the operation and design of their services, as well as their features and functionalities", step e) is stated as a good practice rather than mandatory.

Nine Actions Feedback

Question 2: Do you have any comments on the nine proposed actions?



- Action 1: Policies (pages 22-24)
- Action 2: Risk Assessments (pages 24-25)

Both of these recommend training and risk assessors but there is little detail on how these trainers and assessors are regulated and quality-assured.

- Action 3: Data transparency (pages 26-27)

This does not mention what data collection frameworks are deemed good practice, and how these will be standardised. This part needs to scope how this information can be used in ways falling outside 'data for good' by companies - for instance, there is no mention of how such data (Ofcom is asking platforms to collect) will feed into Governmental policymaking and targets.

- Action 4: Perpetration prevention

Whilst this recognises that methods of misuse are diverse and unique to each platform, this does not mention how an exhaustive OGBV typology will be applied - where the focus often is, understandably, on image-based harms, there needs to be specialist understandings on how more unusual and/or harder-to-detect forms operate, and greater transparency around risk and limitations: what types of perpetration are more likely to 'slip through the net' and why. For these harder-to-reach forms, existent publications should be drawn on - for example, there is a body of research and policy on gender and terrorism prevention.

In relation to harm prevention, YP recommended a differential approach to posters/perpetrators, such as a young boy versus Andrew Tate, thus recognising in some

cases early intervention could prevent radicalisation for bystanders and posters. YP were also acutely aware of measures taken to avoid algorithms.

- Action 5: Safety features (pages 28-30)

This section needs to mention how users will be 'trained up' in such features through awareness-raising, and how accessibility and take-up will be monitored. Namely, this needs to scope limitations, which could form part of risk assessments, by recognising who will be vulnerable to low(er) engagement and understanding of self-protective functions. Currently, the Guidance states user protection in 'flat' ways.

- Action 6: Reduce the circulation of online gender-based harms (pages 30-33)

The Guidance offers up significant opportunities for change, considering the leverage Ofcom possesses and the topic the Guidance addresses, with harmful online content being a societal and policy priority. This compelling combination - of Ofcom's sector position and the timely Guidance focus - is arguably not reflected in the low-strength statement: "it is up to services to decide which methods will be most appropriate in each case", when poor standardisation across firms is the key object of dispute and the Guidance has been issued on this basis. In all OGBV cases, platforms' terms of service use apply.

In addition, although the Guidance's focus on user protections is welcomed, the Guidance could make stronger statements around the sheer volume of, and user exposure to, harmful content. The fact that harmful content is reaching users at unacceptable levels needs to remain front-and-centre, where firms need to recognise they ultimately host this content, user protections are one element of multi-pronged responses, and prevention and removal remain the priority. Harm-reduction is the golden thread running through all elements of the Guidance. This simple 'North star' should characterise, and be consistently returned to, in all recommendations and responses to OGBV.

Finally, we observe Ofcom commits to publishing the adoption standings of platforms 18 months after Guidance finalisation, and to identify non-compliance. Whilst this creates public pressure and enables informed user choice, OGBV will continue during this period and thus, a more crystalised roadmap and theory of change are needed to assure firms will get from 'here to there', pushing what their first targets should be.

- Action 7: Give users better control of their own experiences (pages 33-34)

Action 7 consists of expectations that firms will undertake the steps for user control. Platforms already possess a variety of features and could adopt more but this may take time.

This Action should promote the existent and/or emerging external tools allowing users to impose controls *in combination* with functions from platforms. This multi-pronged combination should be promoted clearer in the Guidance as the 'meta-solution'.

Furthermore, this Action needs to demand visibility and awareness-raising of embedded functionalities to increase user fluency, noting the difference between existence and take-up. Accessibility and utilisation (of provisions) should be emphasised as the superior measures of OGBV prevention, intervention, and support.

We note that users' understanding of safety, and what will help them to be safe, changes over time as OGBV trends do. We recommend the Guidance includes the need for service providers to be longitudinally influenced by service users - here, we mainly refer to YP and victim/survivors as those with the most up-close understandings of OGBV trends.

- Action 8: Enable users who experience online gender-based harms to make reports (pages 34-36)

The recommendation for platforms to allow for "the ability to report off-service abuse [to] enable providers to recognise and address how online gender-based harms are often part of wider patterns of behaviour" needs to tie in with the (expected) capabilities of platforms to capture and intervene with patterned, repeated, and/or sustained behaviour on their platforms.

Furthermore, whilst encouraging reporting is a good measure, little is known about attrition. Data on success rates is unknown. Many reports are not upheld. Platforms need to make this data more available and be given stricter targets. Platforms make their own judgements as to whether incidents violate policies, with little quality control. No oversight is given on platforms' adherence to their own policies - this begs the question of what evidence (of successful adherence) would be sufficient to make claims about meeting targets, and how Ofcom would investigate a failure in report handling where a platform has said no violation has occurred. Action 8 also does not mention appeal capabilities for users.

- Action 9: Take appropriate action when online gender-based harms occur (pages 36-37)

This section is dedicated to the 'aftermath' of OGBV. It does not mention multi-partnership working. It also does not emphasise the more sophisticated ways offenders operate, such as setting up multiple accounts to abuse. It does not discuss how, despite repercussions being in place, trauma has already occurred - this is where prevention, as the most maximal form of harm-reduction, should always be reiterated as the first instance and as the main focus of regulators and of firms. It also does not discuss steps firms should take to signpost support and information for victims and bystanders.

Other Consultation Questions

Question 2 has been answered above. Questions 1, 3, 4, and 5 will be answered in this section.

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

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

Currently, there are four domains of OGBV that the Guidance demarcates. It separates VAWG from other content affecting women and girls but that does not necessarily constitute VAWG. We recommend a more systematic approach, considering a multi-query, multi-dimensional, multi-spatial, pluralistic approach inclusive of Protected characteristics and contemporary configurations of OGBV. We provide the latest iterations of the ESO project's taxonomy and typology as examples.

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 Chapters 3, 4 and 5? Are there any additional examples of good practices we should consider?

We are not commenting on the Foundational Steps as input is only invited to the good practice steps. In the 'Good practice steps: How can service providers go further?', step 3.13 a) on platforms' authoring of policies, there is no mention of Ofcom - or alternative experts such as VAWG charities - playing a role in providing templates and best practice-sharing opportunities. This section needs to stress going beyond firms simply 'having a stance' and stating, in these policies, what they will practically do. In d) on training staff, there is no mention of quality assurance in the design and delivery of this training, nor about the dimensions of staff workers' rights, contracts, and involvement of unions in being trained on OGBV. Reducing abuse and violence should be largely focused on service users, as the Guidance does, but also needs to include trauma-informed, ethical practices for those handling it and/or experiencing it themselves. Thus, as a multi-stakeholder project, ESO encourages the other good practice steps to tie into one another, namely c) consulting with subject matter experts. Point 3.19 a) needs to make concrete suggestions on those able to collaborate and consult in ways that innovate frameworks and concretely hold firms to account for their policy adherence. Ofcom could play a role in facilitating quality external assessors and a system for this. Point 3.26 a) needs to state how data-sharing is done in practical ways to inform policy and strategy, at organisational levels, within Ofcom, and possibly at higher-up governmental levels.

3.26 b) on "Providing more detail about which posts are flagged by automated content moderation, active bystanders who are not targeted by abuse but report content to support others, and the targeted users themselves" needs to detail what can be done with this data. Following standard machine learning practices also adopted in the ESO project, flagged content can be used to train further models to detect content more accurately, especially in the cases of new trends, linguistic tropes, memes, topics, and targets of OGBV based on recent events.

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? X

In the ESO project, we adopted both an 'inductive' and 'deductive' approach whereby our taxonomy framework drew both on contributions 'on-the-ground' from internet users who are OGBV victims/survivors as well as 'top-down' considerations of law and policy. Currently, the Ofcom guidance focuses on user needs, but it does not explicitly reference the UK Government's VAWG Strategy or the Labour Government's target to reduce VAWG by 50%.

Question 5: Do you have any comments on our impact assessment, rights assessment, or equality impact assessment?

In order for A2.5 to marry up with successful compliance, the Guidance should bring in evidence-based understandings of how much effective GBV prevention, intervention, and support - including proper risk assessments - saves private companies money. It is more costly to handle GBV poorly. The table on pages 47-48 does not include such a column; however, the likes of A2.11 go on to state how VAWG causes reductions in users, which is against the financial interests of social media platforms. There is a well-established economic argument for equality, diversity, and inclusion in the private sector that needs to be brought in. Solely focusing on costs may hinder compliance. The business case for tackling OGBV effectively as a worthwhile investment and as long-term cost-saving needs to be properly delivered in the Guidance.

A2.7 mentions "in accordance with the law" but does not comprehensively set these out - firms are not just subject to the Online Safety Act and Human Rights legislation. A2.21 needs a more expansive understanding of adult vulnerabilities within safeguarding.

Case Study

This final section sets out some ESO preliminary findings, where Ofcom's consultation encourages examples, data, and reports from contributors.

Our findings show that with decreasing moderation, OGBV content is prevalent on social media platforms such as Twitter/X, and that it is harder for automated systems to identify implicit and more subtle instances of OGBV. Such systems depend on vast amounts of human labelled data – where people manually review and categorise examples of content to 'teach' automated systems what constitutes harmful behaviour. Our research (Jiang et al, 2024) shows that the individual beliefs and attitudes of the human evaluators makes an impact on their propensity to label content as sexist; thereby affecting the accuracy of the content moderation algorithm. This means that *biased human reviewers can create biased automated systems*. For example, people with attitudes associated with hostile sexism (La Macchia and Radke, 2020; Chulvi et al., 2023) are less likely to label items as sexist, possibly due to the text aligning with internalised beliefs. This highlights a need for stakeholders and

those with lived experiences to have a bigger participatory role in how content moderation systems are created.

Please visit our website to review our list of publications. We encourage the Guidance to create connections between firms and the user-created artefacts available the field - in our case, this would be ESO's taxonomy, typology, and Support Buddy outputs - creating practical change/outcomes through existent user-led research.