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

Refuge's response to Ofcom's Draft VAWG Guidance questions 2 and 3 (detailed response to actions and good practice measures)

May 2025

	Objective 2 Setting out what service providers can do to improve women and girls' safety governance and accountability, testing and service design, and operations and maintenance
	Question 2: Do you have any comments on the nine proposed actions? Please provide evidence to support your answer.
	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.

Blue row indicates an additional good practice step recommended by Refuge.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
Taking Responsibility			
Action 1: Ensure governance and accountability processes address women and		*Foundational step: Board review, accountable individual, written statements of responsibilities, internal monitoring and assurance*	Foundational steps state that an individual on the Board should be held accountable. Although Refuge understands that the foundational steps are not up for consultation, we emphasise the importance that the responsibility for tackling VAWG should never just be one person, who can be replaced or move on, it should be a corporate responsibility which lies with the whole Board. Refuge recommends that Case Study 1 should be amended to be clear that Boards should have collective responsibility for Online VAWG

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
girls' online safety		*function, monitoring trends, codes of conduct, terms of service and PAS, compliance training*.	Refuge recommends that each service provider draws up and regularly reviews an accountability map for the organisation which clearly lays out roles and responsibilities for tackling VAWG. We suggest that this is added to Case Study 1. Refuge believes that the guidance should emphasise throughout that tech companies should take down harmful online VAWG content whilst it is being investigated. The onus should be on service providers to act promptly and with due consideration to safety concerns; and to build in content moderation which takes down concerning content whilst it is examined in more detail. An example of how this could be improved is point 1.12 in the guidance, which states: "Identifying such a broad range of content and activity which is relevant to the experiences of women and girls online does not mean that we expect all such content to be taken down or heavily policed. However, we consider it is right for providers to take a holistic view of the experiences of women and girls online when making decisions about their policies, tools, and features they offer users, and how those choices may impact women and girls' safety." Refuge recommends that 1.12 should read as follows: *"We consider it is right for providers to take a holistic view of the experiences of women and girls online. When you make decisions about your policies, tools, and the features you offer users, you should assess the impact they may have on women and girls' safety and ability to operate freely online. We expect harmful content to be taken down quickly and for providers to set up security systems to ensure safety on and equitable access to their systems. "*

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			Ofcom should produce recommended governance metrics for tech companies so that they can report on action taken in relation to the nine action areas on gender-based harms. The benefit is that monitoring and analysis across the industry would be facilitated. To avoid this becoming a tickbox, Refuge believes that regulation works best if it is a dynamic two-way process. For example, see our recommendation for a 3.26 (d) new good practice step in which companies are encouraged to notify Ofcom of a new or non-designated risk. Refuge recommends Ofcom add good practice steps which address advertising and paid-for promotions of harmful gender-based harms content. Refuge recommends that a good practice step is added to encourage companies to explicitly reference domestic abuse in their community guidelines [and terms of service], which set out a platform's rules for what content can and cannot be posted.
	3.13(a)	Set policies that are designed to tackle forms of online gender-based harms that are prevalent on the service	This note in the Guidance provides examples of 3 specific harms. Refuge recommends that Domestic Abuse is added to this list. Refuge recommends the following to be added to the guidance: Forms of gendered harm such as domestic abuse, stalking, harassment, and intimate image abuse.
	3.13(b)	Ensure that governance and decision-making consider	Refuge recommend that Ofcom add a Case Study which illustrates how tech companies can governance policies which address intersectional online gender-based harms to make it clear to tech companies how this can be implemented

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		intersectionality of online harms	
	3.13(c)	Consult with subject matter experts, particularly those with experience of supporting survivors of gender-based harms, when setting policies and terms of service	VAWG expertise must be properly embedded and appropriately compensated. We have set out how this can be achieved in our response to question 1.
	3.13(d)	Train staff involved in setting policies or governance and decision making on online gender-based harms and safety-by-design	Specialist training is vital. Training should be developed in partnership with the specialist VAWG sector and should always be trauma-informed and culturally aware. Refuge recommends that training in online gender-based harms and safety-by-design forms part of mandatory induction for every senior management and Board appointment and that this point is added to 3.13 (d) in the guidance. This action will mitigate against issues arising from board and staff turnover; it also recognises the importance of responsibility lying with the company as a whole rather than one individual. See also Case Study 2.
	3.13(e)	Create a media literacy-by-design policy to promote critical and informed service use	Refuge agrees with this good practice step. Ofcom's The Best Practice Principles for Media Literacy by Design are well-written and clear. We recommend that this step is strengthened by adding create a *'gender-based harms-informed'* media literacy-by-design policy

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
	3.13(f)	Establish an oversight mechanism for trust and safety decisions	Refuge welcomes this good practice step. We recommend that Case Study 3 include explicit mention of independent VAWG Expertise to be employed alongside other experts
	3.13(g)	Establish an internal accountability process for tackling gender-based harms	Refuge recommends that this good practice step is added to highlight the importance for responsibility to be taken at the very top of the company. This could be combined with step 3.13(a), if it is made explicit about where accountability lies. This step should refer to and enhance Foundation Step 3.11(a) Board review. Set policies, including metrics, through a Board-owned implementation plan, which is monitored annually or when internal or external triggers prompt action e.g. change in product and services or spike in user behaviour or reports We refer you to our recommendation for an accountability map made in Action1 above.
Action 2: Conduct risk assessments that focus on harms to women and girls		*Foundational step: Risk Assessment Internal content and search moderation policies*	We note you refer to Slupska and Tanczer's research on Threat Modelling IPV (in footnote 96). We support their suggestion to design a dedicated "IPV Threat Model" to explore and document avenues for harming IPV victims/ survivors. [1] The distinction made by Slupska et all between threat, harm, and risk is especially useful in achieving clarity about when and how action can be taken by tech companies to reduce and manage online IPV. We urge inclusion of the definition of these concepts in the guidance and for the structure of this chapter to be changed to reflect these definitions.

[1] Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things Julia Slupska and Leonie Maria Tanczer in The Emerald International Handbook of Technology-Facilitated Violence and Abuse, 663–688 Copyright © 2021 Julia Slupska and Leonie Maria Tanczer Published by Emerald Publishing Limited. CH040-9781839828492 663..688

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			Refuge's tech team expertise is ‘state-of-the art’ in how to conduct and manage effective Domestic Abuse risk assessments, sometimes referred to as threat modelling processes. Our support team find that all survivors are different and respond differently with different needs at different times. For example, if they are thinking of leaving, have left or have returned; have children; and so on. And, although underlying patterns of behaviour across perpetrators can be identified, they also behave differently and creatively to adapt to blocks and security put in their way. Therefore, there is a need to regularly refresh risk assessments so that they are up to date. Refuge recommends a quarterly review cycle of risk assessments Refuge recommends a threat modelling and risk assessment approach, which centres women and girls’ online safety, and which takes into account the needs of the most marginalised at the outset (rather than as an add-on). We recognise foundational step 3.17a) iii) talks about 'decide measures, implement and record.', Refuge would like to see risk mitigation policies and implementation plans highlighted in this guidance as they are the necessary corollary to risk assessments, as mentioned in 4.21. This needs to be highlighted more frequently in the guidance to ensure that providers take action to reduce risks that have been identified. Case study 4 gives two examples of how service providers could account for the risks. The second example about online harassment should be amended to mention domestic abuse as a specific risk. Refuge recommends the following addition to Case Study 4

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			> Online harassment can form part of domestic abuse and can include setting up hashtags naming an ex-partner; or sending pictures of their front door. *Refuge's Tech Team shared several examples of online VAWG which could be form the basis of further case studies to help tech companies develop thorough risk assessments* • *Survivors experience harassment which is disguised. A perpetrator may share and harass the survivors with images and videos of clowns on TikTok as they know the survivors would find this triggering but would not be picked up as harmful by the service provider.* • *A survivor who has an Only fans account with image which she had uploaded. The perpetrator has taken images from her Only fans account to impersonate her and created fake profiles on X. They then used these profiles to extort and catfish others on X. This resulted in the survivor getting harassed and threatened online on her real Only fans account.* • *Another example is when a survivor is using a business Instagram account for their livelihood. If a perpetrator takes over the account, bombards it with negative comments and tries to damage the survivor's reputation then this can have a big impact on her livelihood and income.*
	3.19 (z)	Develop tools for risk assessing danger of cyber-stalking or stalking related online behaviour and raise red flag to human	New good practice step: Refuge recommends that stalking is added in as a new good practice step (see response to question 1). Stalking or cyber-stalking is named multiple times in the case studies throughout the guidance, yet it is not named as a distinct harm with corresponding foundational and good practice steps

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		moderation teams to investigate as a priority	
	3.19(a)	Use external assessors for monitoring the threat landscape, including local partners with regional and cultural knowledge, and international partners with expertise in highly contextual risk areas such as cyberstalking and controlling or coercive behaviour	Refuge supports the inclusion of this measure
	3.19(b)	Engage with survivors and victims' to better understand their experiences	Refuge agrees with this good practice step but believes that it should be strengthened. We recommend that the guidance wording is changed to say that it is essential (rather than '*valuable') for* survivors and organisations representing them to share their experiences with tech companies to inform risk levels. However, tech companies need to be sure that they are engaging with survivors in a safe and trauma-informed way, which is usually best mediated by VAWG organisation with well-developed survivor support systems. Refuge recommends that *'trauma-informed'* is added as follows: *Engage with survivors and victims in a trauma-informed way to better understand their experiences.*

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
	3.19(c)	Conduct user surveys to better understand users' preferences and experiences of risk	Refuge recommends that 'co-designed' and trauma-informed' are added because surveys can expose victims and survivors to further harm unless these principles are adhered to. Co-designed research can bring valuable knowledge and first-hand expertise to the research design. This is mentioned in the guidance in the form of a Case Study (5), but Refuge believes this should be brought to the fore as follows: Recommend: Conduct co-designed and *trauma-informed* user *research* and surveys to better understand users' preferences and experiences of risk.
	3.19(d)	Conduct an impact assessment alongside other risk assessments to assess impacts on self-expression, freedom from discrimination, and privacy, especially for those with protected characteristics	Refuge recommends that this good practice list is expanded to include survivor safety. Tech companies should be reminded that many survivors are abused both online and offline and that domestic abuse is extremely serious. A woman is killed every 5 days by her current or former partner, and it is estimated that the number of women who die by suicide following domestic abuse is significantly higher.
Action 3: Be transparent about women and girls' online safety		***Foundational step: Categorised services obliged to produce reports***	Refuge strongly supports the inclusion of this foundation step. We point to the finance sector which is making some progress on women and girls' safety. Refuge supports survivors of economic abuse and works closely with some banks as

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			trusted partners. The 2021 Financial Abuse Code[2] is currently under review and may offer some useful comparison to inform this guidance. Refuge is concerned that service providers could raise objections to providing information based on perceived commercial risks related to public disclosure. Refuge urges Ofcom to strong resist any attempts to water down reports on grounds of commercial confidentiality.
	3.26(a)	Share information about the prevalence of different forms of online gender-based harms and the effectiveness of measures in place to address them	Refuge supports this good practice step. We refer you to our recommended new good practice step below 3.26(d) that this information be reviewed and shared with an advisory group.
	3.26(b)	Provide more detail about posts flagged by automated content moderation, active bystanders who are not targeted by abuse but report content to support others, and the targeted users themselves	We support this measure

[2] Financial-Abuse-Code-2021_Updated_2022.pdf Accessed 18th April 2025

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
	3.26(c)	Exercise caution in sharing information that perpetrators could exploit to circumvent safety measures, as well as details of specific incidents that could identify an individual or group, including location, sexual orientation, religion, or other sensitive information that could put them at risk	Domestic Abuse is perpetrated is many ways. Perpetrators are highly creative at finding new means and ways of locating and abusing their victims. It is particularly important that survivors cannot be identified through this transparency reporting. We recommend that domestic abuse is explicitly mentioned in Guidance point 3.26(c). Refuge recommends that VAWG expertise is deployed to review the transparency process and to dip sample reports prior to public release to ensure that no survivor risks being harmed as a result.
	3.26(d)	Notify Ofcom of new or non-designated risks including the kind of content identified and the prevalence of the content.	To be relevant to emerging risks, Refuge recommends that this good practice step is added to encourage tech companies to alert each other and Ofcom to new forms of harmful behaviour and content so that they can take prompt action. Refuge also recommends (ref Q1 above) that Ofcom facilitate this notification process by creating a chairing a dedicated advisory group, including tech companies, VAWG representatives and Ofcom.

Preventing Harm

Throughout this chapter, Refuge recommends that in addition to a focus on women and girls with protected characteristics, more emphasis and explanation is given about the accessibility needs of diverse groups of survivors. This applies to abusability evaluations but particularly to

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
setting safer defaults good practice steps laid out in Action 5, where it will be important to provide information in a variety of formats, including British Sign Language, using simple language and with easily accessible translate options.			
Action 4: Conduct abusability evaluations and product testing		*Foundational step: Product testing Significant change risk assessment Recommender system testing*	With reference to the abusability evaluations and product testing as a whole, Refuge supports transparent and well-documented upstream safety by design. In addition, regulated services should consider the need for explainability or interpretability, accountability, and auditability in designing AI and machine learning systems, particularly with regard to the representation of women and girls, especially those from minority groups, in their data sets. This action will be even more effective if tech companies adopt cross-company policies and practices which are actively anti-VAWG (see Action 1 above). Furthermore, we recommend co-location of VAWG experts in their teams to advise and develop e.g. abusive and survivor user personas, to advise on abusability testing, and to enhance the application of system testing with VAWG knowledge and experience.
	4.20(a)	Use red teaming for abusability testing	Refuge wishes to highlight the 'people you might know' section or 'suggested people to follow' functionality, which is found on most big social platforms. This can reveal a domestic abuse survivor's new profile to the perpetrator as they may have friends in common, be in a similar area, which enables them to be found if they have moved to a refuge or new location to escape abuse. This functionality should be made optional and not turned on by default, so users have more control over this feature and decide if they want their profiles to be suggested to other users or not. Before activating this functionality, a clear explanation of what this means should be displayed for the user to make an informed decision on whether they want this function or not.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			We recommend that testing procedures and protocols incorporate known dangers for women and girls and are regularly updated for new and evolving risk behaviour. Refuge recommends that search providers are explicitly mentioned in this point, as their content algorithms are often used to promote content harmful to women and girls. Auto-play and autocompletes can often promote misogynist content without the user having specifically searched or looked for it. The GenAI suggestion algorithms provided by search providers should also be safety-informed and abusability tested to ensure that harmful content is not generated.
	4.20(b)	Work with experts with direct or relevant experience engaging with and understanding perpetrator behaviours	See point about VAWG expertise point in response to Q1. As highlighted above at 4.20(a) that perpetrators adapt quickly to evade safety measures. Therefore, the engagement of experts needs to be regular (quarterly and ad hoc to respond to new threats to ensure that tech companies are up to date with evolving threats. Refuge recommends that the following should be added at the end *'on a regular and frequent basis'*.
	4.20(c)	Use personas to explore how different users may experience a feature	Refuge believes that it is essential for the guidance to be more explicit about what is meant by 'different' users to make it clear to service providers about the different intersectional experiences as well as the specific types of gender-based harms including domestic abuse.
	4.20(d)	Adhere to the principles on monitoring and evaluating features in the Best Practice	Refuge welcomes Ofcom's BPPs for MLD. However, it is unclear to Refuge about what specific additional steps are being suggested here. We, therefore, we recommend that Ofcom review and republish the Best Practice Design Principles in 2026, following publication of this VAWG Guidance, to incorporate the main principles.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		Design Principles for Media Literacy	
Action 5: Set safer defaults		*Foundational steps:* *Safe settings* *Group chats* *Supportive information* *Safe search* *Signposting children to support*	Refuge is concerned that Action 5 good practice steps tip the balance towards the user acting rather than focussing on tech company actions. We offer Ofcom suggestions for improvements below. Refuge welcomes setting safer defaults as a positive step. Through our frontline work with survivors, we have a rich evidence base about the differing ability of users to understand what they are signing up to and therefore we recommend that Ofcom add to the Action 5 points (4.22-4.25) about accessibility of options. See also general point made under 'Preventing Harm' above. Multi-factor OR two-factor authentication is only as strong as that method of authentication. Refuge's Tech Abuse team offer the following insight, and we recommend that this is included as a case study. • If it is biometrics on a device, a perpetrator may have access to the device and coerce the survivor to input their biometrics. • If it is voice authenticated, there is software that can deepfake your voice very well. If it is to a phone number, the perpetrator may own that contract. The Refuge Tech Abuse team have had cases where the man who perpetrates the abuse 'owns' (are named with the provider) phone contracts. They report the client's phone as missing. They can then get a replacement sim or a PAC code and take the survivor's number from them. If this number is their 2FA or MFA method, it can be taken by the perpetrator and access to the account can be granted. Similarly with emails used to authenticate, perpetrators may have coerced a password out of a survivor, or they may

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			have kept access to an email account they set up for the survivor. It may not be safe for survivors to remove perpetrator access to their phone, contracts, or emails, whilst in an abusive relationship so platforms should be aware of this. Refuge recommends that providers should make it very clear to users about how to check authentication methods are safe before they are linked to a platform. We also recommend that providers make it easy to report when these methods have been compromised.
	4.29(a)	Set strong and customisable defaults around user interaction	Refuge is in favour of friction mechanisms being introduced when users sign up to new platforms, which will encourage informed choices about defaults. For example, this could involve offering a choice on WhatsApp about when and if to notify contacts and/or group chats that you have changed your number. Refuge clients report that this can become an issue when they chang their number for security reasons and the group is notified which alerts the perpetrator to the change.
	4.29(b)	Set strong and customisable defaults around user privacy	As for 4.29(a) For example, Snapchat's Snap Map is very accurate in its pin pointing of location and should a survivor have a mutual friend of the perpetrators on there, it may make it easy for that information to be accidentally passed back to him, or be added by a fake profile owned by the perpetrator. This and other similar services are a live tracking risk when that app is open on a mobile phone. Refuge recommends stronger guidance to tech companies about automated notifications being sent, if a user changes their defaults, as this could alert a perpetrator and prompt harmful behaviour on other platforms or offline – see 4.29(a).
	4.29(c)	Combine relevant safety and privacy settings into 'bundles'	Refuge agrees with the point in guidance that users must be offered options to make more granular manual choices outside the bundles. This enables domestic abuse survivors to have more agency about their online use, which will enable them to share information and messages with people of their choice.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
	4.29(d)	Strengthen account security with two- or multi-factor authentication feature	Refuge welcomes this but also cautions that account security settings need particular attention where domestic abuse has been identified either by the user concerned or by patterns of behaviour. Refuge has had many reported cases of perpetrators of domestic using tech companies' reporting systems to control a survivor (see Action 9). See point under Action 5 about perpetrators getting multi-factor authentication sent to their own devices. Refuge is willing to offer a case study to Ofcom on 'how to support users to secure their accounts' illustrating the harmful circumstances, how the trusted partner status worked, and the good practice steps the tech company took.
	4.29(e)	Provide information about account access	The guidance on this says: "e) Account access: Providing information about account access by making it clear which users are currently connected to an account, device or platform, as well as what unique delices (via IP/MAC addresses) are connected to an account. This minimises opportunities for non-consensual monitoring and surveillance." In Refuge's view the guidance needs to be strengthened here so that tech companies are clear about the importance of providing evidence to survivors and relevant law enforcement agencies, about the use of technology which is causing harm. This is crucial both to: A) be able to evidence to police the misuse of technology, as, unfortunately, often the onus is on the survivor to show police that they can source 'enough' evidence of abuse before police officers will begin an investigation

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			B) It will support survivors and any support worker they are working with to understand how a perpetrator may be accessing information about them or their children Refuge recommends that an additional point 4.28a should be added (between existing 4.28 and 4.29) to explain to tech companies about the importance of providing account access information to survivors, and, with relevant law enforcement agencies. In our response to question 1, we set out in further detail how the guidance should be strengthened to create clear protocols with sharing information about the abuse perpetrated on platforms with survivors and law enforcement agencies.
	4.29(f)	Identify optimal frequency and timing and give users regular reminders for reviewing or updating privacy and security settings	Refuge agrees with this good practice step and recommends that the following is added: "*with easy-to-understand language with options for translation*"
Action 6: Reduce the circulation of content depicting, promoting, or encouraging		*Foundational steps: Automated content moderation* *Recommender systems* *Search moderation* *CSAM warnings for search*	Refuge notes that there is an absence of guidance about the measure's tech companies should take to co-operate and facilitate the prosecution of criminal and civil cases, where request to do so by the survivor and/or by the appropriate authorities. As set out in our response to question 1, we recommend the guidance is strengthened in this area. *Refuge welcomes action 6 but recommends hybrid moderation (automated and human moderation) as best practice. It is also crucial that the speed of moderation and need to remove online VAWG content quickly is highlighted.*

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
online gender-based harm		*Highly effective age assurance*	*Example from a workshop with the Refuge tech abuse team:* *When an image of a woman is posted online, a screenshot takes a second and lasts a lifetime. We have had cases of survivors who have been sent screenshots of their images, or of posts about them from worried friends or family. Then when they report the post to the company, the man who has perpetrated the abuse has already taken it down, so no action is taken. As time passes, it is likely more copies of the content have been made.* Refuge recommends that the Guidance should add to introductory section an additional point as follows: The mix of approaches recommended in 4.32 need to be monitored and assessed to ensure that a whole picture of individual's behaviour is available to assessors and moderators. This is particularly important to protect survivors of domestic abuse (and victims of stalking) from perpetrators who use a variety of methods to harass and abuse their victims.
Persuasion	4.40(a)	Introduce deliberate friction through nudges at the point of upload	Refuge agrees with the introduction of friction when users use Share Buttons to spread misogynist or other VAWG content. Share buttons allow the rapid sharing of content from one platform to another, which for many people is positively used, but in incidents of doxing or intimate images, it allows the abuse to move off site very quickly making it much harder for survivors to report to all social media platforms. However, we note that nudges are likely to have little to no effect on perpetrators of domestic abuse, where they intentionally want to harm an individual and therefore will ignore this type of message.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			As per our response to question 1 and reiterated above, it is vital that the guidance is strengthened to include the provision of data to support survivors or law enforcement agencies pursuing criminal or civil cases. This data should include the friction measures or nudges that were in place and seen by the perpetrator.
	4.40(b)	Allow users to verify their identity	Full text in guidance as follows: Perpetrators of online gender-based harms often use multiple accounts, including anonymous accounts to perpetrate abuse against survivors. Example from Refuge's tech abuse team: *A dating site, where users are all verified, removed a user (perpetrator) following a report of domestic abuse. The abuse on this platform was stopped and the user (survivor) remains confident that the perpetrator will not get back onto the platform, as the process for account set up requires verification. This did not stop the man going on to use other dating other sites, however, where these processes are not in place.* However, regarding verification is important that survivors can be online anonymously for their own safety. For example, as part of safety planning, a caseworker might advise a survivor to not use a profile picture featuring her face; not use her full or real name. However, with verification this may be harder. We recommend that Ofcom work with tech companies and VAWG organisations to find the right balance between perpetrator accountability and survivor anonymity. One area to explore is platforms requiring verification, but this data is not made available on public profiles.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
Removal	4.41(a)	Use hash matching to prevent uploads of known intimate image abuse	Refuge welcomes the inclusion of this important safety measure
	4.41(b)	Implement timeout features to users who repeatedly attempt to abuse a service to perpetrate online gender-based harms	Refuge welcomes the inclusion of this measure, but notes it is likely to be ineffective in the majority of cases of domestic abuse due to the motivation of the perpetrator to abuse a specific survivor as part of a pattern of coercion and control. However, it could be an effective friction measure for some other forms of online VAWG The Refuge Tech Abuse team have highlighted cases where, if a cooling off period is enforced, the perpetrator will move to different platforms to continue abuse. An example is messaging via payment apps such as PayPal and making abusive remarks in reference lines. Once cooling off or temporary bans expire, the abuse continues again. Survivors report being 'exhausted' by the reporting process and feel that they have no option but to reduce their online presence. This compounds their isolation and can impair their ability to stay safe offline. We highlight our points made under Action 8 to make reporting easier, which, in turn, will lead to people who perpetrate abuse being reported more often, which should result in more perpetrators removed from platforms. The guidance should also recommend that platforms act with stronger deterrents if timeouts are repeatedly ignored, for example platform bans.
	4.41(c)	Require consent from those depicted in intimate content prior	Refuge's concern here is that women and girls can be coerced into giving consent, so to be effective, this action, along with ID verification, needs to form part of comprehensive approach to tackling controlling and coercive behaviour targeted at women and girls. User

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		to uploading where adult content is allowed on a service to prevent intimate image abuse, including deepfakes	requests to take down pictures, where consent has previously been given (whether voluntarily or not) need to be dealt with promptly and the person making the request must be believed. See our comments about the sensitivity required to provide safe reporting for survivors of domestic abuse. in Action 9 below.
	4.41(d)	Implement prompt and output filters for GenAI models	Refuge supports this measure. Gender-based harmful content is being included in GenAI results and models. This is particularly concerning when survivors are searching for support. In testing AI models in March and April 2025, Refuge has generated inappropriate answers or tone in AI generated answers, when we pose as survivors seeking help, empowerment, and support. We recommend the guidance contains more information on the importance of developing safe GenAI models. These models rely on existing content to generate answers, which may simply amplify the harmful content the guidance is addressing.
Reduction	4.42(a)	Deprioritise harmful content in recommender algorithms to reduce its visibility and reach	Yes
	4.42(?)	Introduce tools to reduce or slow down virality of harmful content	Refuge recommends this good practice step is added because it is at the core of service providers' business model and therefore that they understand how to increase the speed with which content share and know how to reduce it when harmful content is involved.
	4.42(b)	Remove links to sites known to host	Refuge strongly supports this measure.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		nonconsensual images, or to services such as nudification apps	Refuge recommends inclusion of a case study about intimate image concerns in domestic abuse cases. The Refuge tech team provided two examples in a workshop to respond to this consultation which illustrate the importance of this measure. *Example 1: A Refuge client shared images consensually in early days of relationship but has never consented to the sharing of images publicly. The person perpetrating the abuse has threatened to share the pictures with the survivor's family and friends via social media platforms (because the survivor's family and friends have blocked him from direct messaging). Following Refuge's recommendation to use the hash-matching service provided by StopNCII, she has subscribed to StopNCII's hash bank to prevent the images being shared on other platforms. At present, no sharing has happened, but the fear remains.* *Example 2: In honour-based violence cases image circulation can escalate the risk of harm from the abuser and from his family and friends. To note that images which are dangerous for honour-based violence survivors may not be explicitly sexual but could depict survivor in same sex relationship, wearing clothes deemed inappropriate or drinking alcohol which are viewed as transgressing cultural and religious rules or norms. These images also need to be taken seriously and removed quickly as well for survivor safety.*
	4.42(c)	De-monetise sites that promote online-gender based harm	We welcome this measure - see also point made under Action 1 about Governance policies addressing advertising and paid-for promotions of harmful content.
	4.42(d)	Blur nudity and harmful content	We welcome this measure.
	4.42(e)	Scan for duplicates of explicit non-consensual fake	We welcome this measure.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		content and delist them from search	Reverse image search provided by platforms such as Pimeyes is used by survivors to track their images and content across multiple platforms, which helps them with reporting abuse to tech companies. However, Pimeyes has also made it easier for perpetrators to find links to information about the survivor's online and offline activity and location. This is a good example of the need for knowledge and understanding within tech company moderation teams about achieving the right balance between safety and control for women and girls.
Automated Detection	4.43	Use automated content moderation to scan, identify, and filter online gender-based violence content	Refuge notes this good practice step, and we also note that perpetrators of domestic violence often use voice-notes, emojis and other non-text-based content to communicate harmfully with their targets. We recommend that it is strengthened by adding *'Including non-text-based content.'* We recommend a case study is included to illustrate how domestic abuse perpetrators harass and abuse their victims across multiple platforms including the use of voice notes, emojis. This extends to mobile phone services, the misuse of which service is mentioned in Case Study 20 but needs strengthening. We recommend that Case Study 20 is referred to here.
Supporting Women and Girls online			
Action 7: Give users better control over		*Foundational steps: Block and mute, disable, negative*	Refuge is supportive of measures to give greater control over women and girls' online experience as described in the Chapter 5 target outcomes (5.7-5.10).

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
their experiences		*feedback, group chats, supportive info, support materials*	Refuge sets out in our response above the importance of providing accessible information to enable users to make choices and better control over their experience. Although accessibility is mentioned (5.1, 5.6, 5.8), we believe that this point is insufficiently visible in the guidance. Tech companies should be encouraged to routinely develop guidance for users experiencing online VAWG which sets out actions they can take; we also recommend that information about tech abuse should be promoted and be made available to all users, who can report online VAWG as third-party bystanders. We recommend that Ofcom include a reference to Refuge's safety resources in the guidance, which are housed on a dedicated website called Refuge Tech Safety. The resources include a series of step-by-step support guides for a range of devices and social media platforms, and an interactive chatbot with video guides in multiple languages. Given the vast resources available to many social media platforms, these platforms should be able to develop similar resources for their users.
	5.15(a)	Allow users to delete or change visibility settings of content they upload, including content uploaded in the past	Refuge supports this good practice step as it provides survivors of domestic abuse with the option to control their past, current and future digital footprint.
	5.15(b)	Provide users with tools to block and mute multiple accounts simultaneously	Refuge is in favour of this good practice. A common experience for survivors is to receive dozens or even hundreds of abusive, harassing images or communications from the person perpetrating the abuse. Due to the limitations of current reporting processes, survivors must report each individual piece of content in turn, which is both time-consuming and re-traumatising.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			However, tech companies (and users) also need to be aware that blocking and muting can prompt further abuse offline, especially in cases of domestic abuse and stalking. It is also worth noting that for some women and girls, they feel safer if they can see what the perpetrator is posting, rather than blocking them. Service providers need to employ independent third-party VAWG experts, like Refuge, to inform their good practice, as perpetrators' behaviour evolves to exploit current and emerging technologies. Refuge recommends offering users functionality to see what the other user(s) see when they block or mute prior to making the decision. Refuge recommends tech company investment in technologies which enable online platform to recognise when one user has set up multiple accounts, such as by identifying where one IP address has been used to create numerous accounts. *"A website where you could mass block someone across other sites would be the dream"* (Member of the Refuge Survivor Panel).
	5.15(c)	Allow users to filter out content from all users who have not completed identity verification	See point made against 4.40(b) about issues with User ID verification
	5.15(d)	Provide users with greater control over what content is recommended to them	The recommender functionality is deployed by many social platforms. Due to the algorithms of these platforms the survivor's new profile could be revealed to the perpetrator as they may have friends in common or be in a similar area, for example. This safety functionality needs to be explained clearly, and users sent reminders about how to do it.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		by content recommender systems	Refuge recommend that this includes functionality to 'block your own profile,' such as in the 'people you might know' section or 'suggested people to follow.'
	5.15(e)	Allow users to signal what kind of content they do not want to see, and what kind of content they want to see more of	Refuge supports this good practice step. We believe that it is important for users to be informed about search engine functionality that links searches to accounts or profiles, e.g. Google search users should be informed about the algorithm which builds profiles of users based on their search history. There could be danger present for women and girls if they search for support online and that this then is added to their profile, which in turn throws up search suggestions, which could enable a perpetrator to deduce that they have accessed safety content. On the other hand, this functionality can be used positively, for example a survivor finding and engaging in an online domestic abuse support group and this then linking her to a new safe area to get support online.
	5.15(f)	Signpost users to supportive information which addresses specific harms such as domestic abuse or image-based sexual abuse	Refuge welcomes this suggestion. However, we strongly recommend that this point is strengthened to ensure accessibility for all survivors See the point made against Action 7 by way of good example.
Action 8: Enable users who experience online gender-based harms		*Foundational steps: Complaints processes and systems and communications; predictive search*.	Refuge welcomes Ofcom's guidance on reporting and the reference you make to our report 'Unsocial Spaces', which provides evidence about poor experiences with reporting systems and how that erodes survivor trust. In turn, leading to a reduction in the number of reports being made by women and girls. We recommend that an additional point is made here (5.19) to emphasize the importance of reporting and mitigation actions be reported regularly to Board level. This is to ensure

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
to make reports			that the feedback loop between governance and practice is complete so that policies and practices can be updated and amended in line with user experience. Reporting systems also often rely on checkbox systems to indicate why content is harmful, which rarely list domestic abuse, and require users to report individual pieces of content, which can be retraumatising and time-consuming for many survivors, as perpetrators often send dozens or even hundreds of abusive messages. We refer you to our points on about a) multiple blocking b) sharing abusive users' IPs across platforms and c) improving reporting processes.
	5.20(a)	Provide a 'quick exit button' throughout the reporting process which immediately takes the user out of the reporting system	Refuge agrees with this functionality being encouraged. However, it is important to note *that most 'quick exit' functions do not delete browsing history. If you are accessing sensitive materials from a shared device on a 'standard' browsers like Chrome, Edge, etc, a list of addresses visited is automatically saved to the browser. If a perpetrator suspects a survivor is seeking support, they could check the history, and despite the woman 'quick exiting' a site and not being 'caught live' on the content, the history could give away actions*. Tech companies should be encouraged to provide a guide to deleting browser history as well as quick exit buttons.
	5.20(b)	Allow users to track and manage their reports and tailor their experience throughout	Refuge recommends that survivors are offered the option to use this functionality. However, we encourage the instatement of minimum warnings to users that make it clear that their reports history could be seen by anyone with access to their account.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		the complaints process	For information and by way of explanation about the model used by Refuge, our Tech Abuse team help a survivor to secure their compromised tech in a specific order: 1. Secure the survivor's account and gather evidence 2. Move to formal reporting if/when identified as safe. See also Case Study 21 – where there is an opportunity to add additional good practice steps to track domestic abuse-related activity.
	5.20(c)	Allow users to give feedback to the service provider on their reporting process	Refuge welcomes this measure.
	5.20(d)	Establish a trusted flagger programme in partnership with organisations that have expertise in gender-based harm	Trusted flagger programmes are important and can be effective. However, it is crucial that these programmes are meaningful and lead to a full and swift response. In Refuge's experience, some platforms create trusted flaggers but do not provide an enhanced service, leaving organisations like Refuge waiting extended periods or receiving generic responses – this undermines survivors' confidence in tech companies further. In addition, in the absence of a centrally regulated mechanism to provide expertise to the tech sector, it is important that service providers employ and pay for VAWG expertise in line with their own pay grades for equivalent domain expertise. We refer you also to our comments against Case Study 22 Trusted Flagger.
	5.20(e)	Allow users to report incidents of abuse, including abuse that	Tech companies would benefit from adopting a consistent internal approach to responding to reports and to share that with external sector experts and/or trusted flaggers. Support for survivors is more effective if survivors can be provided with accurate

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		happened on another service or offline	information about how to report, which, in turn empowers the woman or girl to make their own choices about what and how to report. It is also especially important that a survivor can report activity by an individual person rather than having to report every post. See our point under Action 8 above. For example, Refuge's Tech Abuse team report that a co-parenting app has installed a tone checker which recommends changing the tone of speech to something less abusive. This can lead to the person perpetrating the abuse to another less direct way to phrase things and may message the survivor multiple times in an hour about the child's welfare to continue the harassment. On the face of it, each individual message may not constitute abuse, but the impact is traumatic. Refuge recommends the following good practice steps are added to the guidance. Information about how the reporting process works to be shared with users and with trusted flaggers. Updates on reporting progress to be provided to users. • Acknowledgement of a report within 24 hours. Serious offences should be actioned in 24-48 hours maximum, and within 3-4 working days for less serious offences. • Refuge recommends that domestic abuse is listed in the list of 'reasons for reporting' provided by reporting tools. It would benefit users to be given space/options to give a more detailed description about the harm they have experienced and how this is contributing to the domestic abuse being perpetrated against them • Refuge recommends that if domestic abuse is selected, that the user is given the option to report the user or a number of posts in a single report

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			• Human interaction is important when reporting domestic abuse -it is key that survivors can have a point of contact to update their reports as abuse progresses on the platform to prevent continual re-traumatisation
	5.20(f)	Adopt the principles on user-centric design and timely interventions in the Best-Practice Design Principles for Media Literacy	Refuge recommends that this good practice step is strengthened by adding: *'for products and services as well as reporting and complaints process'*
Action 9: Take appropriate action when online gender-based harm occurs		*Foundational steps: Take down, performance targets, prioritisation, moderation teams, complaints, appeals.*	Refuge is highly supportive of this action Refuge urges Ofcom to encourage tech companies to support the adoption of minimum standards, to publish their adherence and report publicly on success in meeting them. Refuge recommends the addition of *'prompt' or 'timely'* to 'appropriate action as speed is often essential in mitigating harm to survivors. We recommend the inclusion of a Case Study to illustrate what a good user journey could look like when reporting online VAWG
	5.25(a)	Take action against users who continually violate a service's terms of service	We recommend that Ofcom clearly states that all service providers have a duty-of-care to users. For example, it is particularly important that is action includes that which taken by service providers, as timely action, which is uninitiated by the survivor, reduces the work a survivor must do to keep herself safe. Refuge recommends the addition of 'prompt' or 'timely' as follows

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			We are concerned that a 'strike' system allows continued bad behaviour (up to three times) which can allow the generation of yet more content and distress for the survivor before the perpetrator is banned. We recommend that in instances where domestic abuse is reported that the account is suspended while investigations take place by a human moderator. Content moderation and search teams must consider the potential for domestic abuse when responding to user reports of harm. Refuge agrees with Ofcom that trusted partnerships are good practice, as suggested in Case study 24, with Refuge and other orgs, to benefit from frontline experience and expertise. Content moderators and search teams need to be aware of the possibility that a perpetrator may report a survivor to a tech company to get her suspended – reducing her access to online platforms and as a way of harassing her and cutting her off from support services, friends, and family. A request for reinstatement may come from a survivor or a specialist VAWG support service. Tech company teams need to appoint human moderators who can assess these cases as a priority.
	5.25(b)	Add fact-checking and labelling to content, which can be a useful tool to address gendered disinformation	Refuge agrees with this good practice step.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
	5.25(c)	Add watermarks and metadata	Refuge recommends that users are prompted when making and posting pictures about watermark and metadata functions. We recommend making this functionality default with opt out.
	5.25(d)	Identify and prevent the creation of new accounts by banned users	Refuge recommends that service providers set up reporting systems which enable rapid identification of the same person abusing the service, and that cases of domestic abuse are prioritised. The Refuge Tech team have found that perpetrators are very easily able to create a new account and continue to abuse the survivors. We believe that companies may be able to make more effective use of IP address information to stop them. There have been cases where women have been harassed on X by the same group of men's rights activists. They were being banned and accounts deleted from X, but they quickly began new accounts to continue the abuse. Refuge recommends that Ofcom encourage tech companies to use IP address information and mobile phone numbers (in addition to email addresses) to identify and prevent abusers from setting up multiple accounts. We urge Ofcom to consider the issues associated with non-UK based IP addresses, and to issue guidance about how to keep women and girls safe from offshore accounts.
	5.25(e)	Send high-risk and highly contextual user reports of gender-based harms for	All content moderation needs to be safe-by-design – both automated and human. All staff who design content moderation systems and who moderate need to attend VAWG training when being inducted and regularly.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
		review by specifically trained moderators	If domestic abuse gender-based harms are identified or reported, then, as footnote (290) suggests, this should be escalated to a specialist *human* service immediately and appropriate mitigating action taken within 48 hours e.g. taking down content; and the user blocked – see 5.25(g) above. Refuge recommends adding the following as a good practice step not just a footnote: (see 5.25(h) below. *Employ human VAWG trained moderators to address high risk and highly contextual reports*. Refuge recommends the introduction of recommended minimum standards for service providers to monitor their activity and progress and to report against to Ofcom and the public in terms of service and community guidelines. Refuge recommends an Ofcom managed VAWG advisory group made up of independent third-party specialists, including domestic abuse specialists, to support the specialist moderators employed by service providers, to review general content, patterns, actions being taken, and advise on what 'else to take into consideration' in order to stay alert to and address the dynamic and rapidly changing ways in which tech can facilitate VAWG.
	5.25(f)	Hide potentially harmful content while it is assessed in content moderation	Refuge agrees with this proposal, which also protects women and girls from the content being shared during the time the assessment is taking place. We refer you to our point against 5.25 (a).
	5.25(g)	Create dedicated reporting and review channels for online gender-based harm	Refuge agrees with this good practice step but believes that these reporting and review channels are only useful if incorporated into the accountability review cycle which goes includes the Board and product design and review cycles.

9 Actions	Guidance Paragraph	Description	Refuge comments and recommendations
			We therefore recommend that Ofcom link this good practice step and others under Action 9 to the 'Taking Responsibility' Actions above. The Refuge Tech Abuse team recommends that tech companies are encouraged to provide survivors with the functionality to report anonymously, and to opt to get real time updates about the state of their report e.g. has it been seen? has action been taken? What action has been taken? This functionality is so helpful for women and girls who are constantly having to make and update safety plans. Refuge recommends that footnote 291 which refers to language needs when reporting, should be expanded to include assistive technology for people with disabilities. See our point above about accessibility. Refuge recommends the following addition to this good practice as follows: Create *accessible* dedicated reporting and review channels for online gender-based harm

Case studies

	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.

Case study No.	Facilitator questions/comments	Tech/Policy comment
TAKING RESPONSIBILITY		
Action 1: Ensure governance and accountability processes address online gender-based harms		
Case study 1: Governance and accountability Ensuring that considerations for online gender-based harms are embedded within the organisation and its systems and processes could include: • A senior person within the service provider being accountable to the most senior governance body for ensuring the service considers and addresses online gender-based risks. • Terms of service and/or community guidelines that are specific and accessible. This should include regular, systemic reviews to ensure that they remain effective, proportionate, and responsive to developing trends in online gender-based harms. • Monitoring and assurance focused on tracking emerging threats, such as deepfakes and other developments in GenAI-enabled abuse, and evaluating whether safety measures adequately address them. Please see Chapter 4 for further details on how products can be tested for resilience to emerging threats		Refuge recommends that this case study is strengthened as follows: Set policies, including metrics, through a Board-owned implementation plan, which is monitored annually (minimum) or when internal or external triggers prompt action e.g. change in product and services or spike in user behaviour or reports. Renew every year, publish on website, and integrate into terms of service and community guidelines. See Refuge's recommendation for an additional good practice step 3.13(g) to highlight the importance and effectiveness of responsibility to be taken at the very top of the company. This could be combined with good practice step 3.13(a), if it is made explicit about where accountability lies. This step should refer to and enhance the Foundation Step 3.11(a) Board review. Refuge recommends that an additional bullet point is added at the top of the case study which states that accountability lies with the Board, to manage a

		dynamic implementation plan, which always ensures accountability – see this point made against Action 1.
Case study 2: Sexualised harassment policy Women, particularly women from ethnic minority backgrounds, can be sexualised and fetishised online in ways which harm their sexual autonomy. 92 This can include unsolicited sexual messages, sexual deepfakes, and images reposted with sexual comments that fetishise or degrade women. Sometimes sexualisation can be implicit or highly contextual. Where a social media provider notices users resharing women's posts in ways which sexualise them without their consent, the company can set out a harassment policy which makes it clear that sexualising someone without their consent is a violation of the policy. Having clear, specific policies on the nuances of online gender-based harms is an important step for creating safer spaces for women online. This is particularly important for nonconsensual sexualisation, which can often be normalised so much that it goes unnoticed		Refuge agrees that having clear specific policies on the nuances of online gender-based harms is a crucial step. Refuge recommends that this case study is strengthened by explaining that perpetrators of domestic abuse are known to abuse and harass survivors in many ways including through sexualised harassment. What might appear to be a 'normal' post can form part of a series of attacks. Therefore, the sexual harassment policy needs to give clear guidelines for staff and the public that the context of the harassment will be addressed as well as the content.
Case study 3: External oversight Service providers' community guidelines and content moderation decisions make them powerful arbiters of public speech. This can have serious implications for users, particularly in cases where biases around characteristics such as gender and race are embedded in service providers' policies, such as content moderation guidelines.		Refuge recommends that Case Study 3 includes explicit mention of external VAWG expertise to be employed alongside other experts. See also good practice step 3.13(f)

For example, content moderation decisions based on an over-zealous application of policies have led to removal of posts about mothers breastfeeding, LGBTQ+ couples kissing, or survivors sharing their experiences of sexual violence. Likewise, algorithms used for content moderation may have biases embedded in their training data, leading to biased outcomes in which women from ethnic minority backgrounds are disproportionately policed for online speech. To introduce accountability and oversight for bias in decisions and policies, companies can engage with external experts or set up an external appeals ombudsman. Such an ombudsman could accept complaints from users appealing content moderation decisions and provide feedback that helps clarify a provider's policies. This process enables service providers to have clearer and more consistent rules for content moderation, leading to safer experiences for women and girls online.		
Action 2: Conduct risk assessments that focus on harms to women and girls		
Case study 4: Gender-sensitive risk assessments • When assessing their risk level for different kinds of harm, service providers need to consider risk factors, including functionalities, business model and user base. • As a starting point, service providers may assess their user base demographics to understand which harms disproportionately impact women and girls, particularly those with intersecting identities.		Case study 4 gives two examples of how service providers could account for the risks. The final paragraph about online harassment needs to mention domestic abuse as a specific (and fourth) risk. Recommends the following wording –

• For example, young women (aged 18-24) are particularly at risk: one in five women in the UK have suffered online abuse or harassment, increasing to one in three for young women aged 18 to 24. 106 Young women are particularly at risk of image-based sexual abuse including cyberflashing and intimate image abuse. 107 In addition, while they are at an especially high risk of online gender-based harm, they fall into a protection gap as many safety measures are aimed at children under 18. • Service providers often collect demographic data about users, for example for advertising purposes or to improve users' experiences. 108 Sometimes services can also make inferences about demographic data on the basis of user behaviour, for example inferring age or gender from the kinds of videos users watch.109 Providers can also use such user data at an aggregate level to consider risks to women and girls on their platform (such as the proportion of their user base that could be at risk). When considering the use of personal information, providers must also consider privacy rights and comply with duties under the UK General Data Protection Regulation ('UK GDPR'). • In addition, services can assess risk with a specific focus on women and girls. Taking the example of online harassment, services can understand how: > User base demographics can show that online harassment disproportionately affects women and girls, and in particular women in public life, as well as women and girls with multiple protected characteristics;		> Online harassment can form part of domestic abuse and can include setting up hashtags naming an ex-partner; or sending pictures of their front door. It is also important that risk assessments can cover business profiles. Online VAWG can also involve small businesses, and their accounts run by survivors. They have had reports from survivors of alleged perpetrators infiltrate business spaces and negatively impact the business economically. They also see issues with 'live streams' which are far harder to monitor and police due to delays in assessing harmful content if only available live and immediately. These may need their own risk assessments due to the unique nature of 'policing' live content. One suggestion is for platforms to use data from text-based comments and report history to review who can go live. Disabling the feature or restricting how widely the live can be shared might be a consideration if moderation cannot happen in real time.

> The functionalities of the service and business model – such as reposts or trending hashtags which amplify virality of hateful content – can contribute to harassment; > Online harassment is often sexualised and includes elements of body-shaming and fetishisation can manifest in ways that overlap with other harms, including offline stalking or threats.		
Case study 5: Trauma-informed user surveys The nature of online risk changes rapidly, as perpetrators of abuse identify new ways to co-opt or subvert new technologies to coerce and harass their targets. Conducting user research such as surveys can be valuable to identify these developments and respond to them. For example, users of dating platforms are particularly at risk of abuse, such as stalking, harassment and coercive control in the context of dating. Chayn, an organisation that supports survivors and victims of domestic abuse, has partnered with various online dating platforms to help them better understand and support the risks their users experience. Chayn helps companies apply principles of trauma-informed design in their user surveys, with a focus on informed consent and privacy, sharing context on how data will be used, and working with a localisation team who understand trauma for multi-language surveys. Such surveys can lead service providers to, for example, develop safety tools to prevent cyberflashing on their service (please see Chapter 4 for further details on preventative measures). The survey also signposted to		Refuge recommends that this case study is changed to *"Trauma-informed research and surveys"to* drive home the importance of research and not just surveys.

relevant resources, including Bloom, a remote trauma service offered by Chayn to online dating users who report harassment, assault or abuse. The service includes courses on healing from sexual trauma as well as access to one-to-one chat support and up to six sessions with a trauma-informed therapist		
PREVENTING HARM		
Action 4: Conduct abusability evaluations and product testing		
Case study 6: Abusability product testing • Abusability testing is one kind of product testing that can inform online safety risk assessments143 by internally testing a product to see if or how it can be abused before deploying it. This testing can also be applied to products which have already been deployed, as not all incidents of abuse will be reported by users and therefore may go unnoticed. • In the context of online gender-based harms, abusability testing involves understanding and mapping out which features and functionalities are likely to be misused for harms such as domestic abuse, harassment, intimate image abuse, and stalking. • Sometimes, these features could also be low risk for other users, or have high levels of utility for some users, creating difficult trade-offs. For example, while some users may abuse location-tracking for stalking and surveillance, others may benefit from being able to monitor their children or relatives. • In some cases, removing such a feature could reduce the use of the application for adversaries without substantially inconveniencing legitimate users. In other		Case study 6 Refuge are in favour of this case study. This type of testing is most likely to be effective if it a collaboration between engineers and developers and VAWG experts to ensure the different ways in which perpetrators abuse are fully understood and taken into account. We also recommend adding a bullet point, which outlines the need to ensure notifications are accessible for all survivors.

cases, the service provider may judge that the feature offers benefits to the majority of its users, but to manage the risk to some users, it could provide better information or customisable defaults. Some features are commonly co-opted for specific forms of online harassment. For example, a feature which allows creating and sharing lists of other users can be misused to share lists of targets with specific characteristics for pile-ons and coordinated harassment. Design changes, like notifying users if they have been added to a list, seeking their permission before they are added, or allowing users to remove themselves from these lists, can mitigate the risk of harm from such features.		
Case study 7: Red teaming for non-consensual intimate image abuse deepfakes The proliferation of audio-visual GenAI tools has facilitated the rise of non-consensual intimate image (NCII) abuse deepfakes, leaving a devastating impact on the lives of survivors and victims, most of whom are women and girls. While many services with GenAI functionalities employ safeguards to prevent the generation of deepfake intimate images (such as safety filters), research shows that users can successfully break guardrails leveraging basic prompting techniques. For example, malicious users have circumvented online tools to generate deepfake intimate images of women in public life. Ongoing red teaming can help service		

providers make their GenAI tools more robust against such attacks. Red teaming is a type of model evaluation that seeks to find and fix vulnerabilities in GenAI models. Service providers have used red teaming to understand whether their model can produce explicit or harmful material. In our discussion paper on red teaming, we considered good practices for a red team exercise, which could involve a service provider testing the effectiveness of safety measures intended to prevent AI-generated intimate images of women in public life. It could involve a service provider testing the effectiveness of safety measures intended to prevent the generation of sexual content. The provider could then test the model's ability to generate images of public figures. If both tests are successful, it could in theory indicate the likelihood of the model being used to generate deepfake intimate images of public figures. In instances where such vulnerabilities are identified, the service provider should take steps to strengthen its existing safety measures. This can include improving its input and output filters (such as content filters), updating blocklists for specific public figures, and removing nudity content from training datasets.		
Action 5: Set safer defaults		
Case study 8: Preventing grooming through safer defaults • Online grooming involves establishing and developing communication with children for the purpose of conducting child abuse. Perpetrators often use bribery,		

blackmail or coercion during grooming. The majority of children targeted by grooming are girls, in part because girls are often seen by perpetrators as being more vulnerable to being targeted. • Perpetrators often target services that encourage new connections, such as social media or gaming services, particularly those with many child users. After establishing communication, perpetrators often attempt to move communication to another service, particularly private messaging services which allow perpetrators access to image-sharing. • Default safety options to the highest protection level can help prevent online grooming, such as disabling the option for unknown adults to contact children (messaging) and hiding information about them (geolocation)		
Case Study 9: Removing geolocation information by default Providers often collect and share information about users' locations. For example, smartphone cameras use information from mobile data, Wi-Fi, GPS networks and Bluetooth and embed this as location metadata in photo and video files. Many users are not aware that when they share photos and videos that include location metadata on social media and messaging services, they can inadvertently reveal users' locations. Likewise, many providers will collect and share users' locations to enhance social networking, which can lead to unintended consequences. Such information leaking can		

cause serious harms, up to and including homicide, in cases of coercive and controlling behaviour and stalking. To prevent such harms, some providers limit opportunities for location sharing, ensure geolocation options are off by default, and provide obvious signs and warnings for users when location tracking is active. For example, services can ensure that metadata is removed from all images upon upload. Policies which ensure privacy settings by default can also be effective in protecting users from online gender-based harms		
Action 6: Reduce the circulation of content depicting, promoting or encouraging online gender-based harms		
Case study 10: Hash matching for CSAM • The circulation of CSAM online is increasing rapidly. Child sexual abuse and the circulation of CSAM online causes significant harm, including to girls, and ongoing circulation of historical imagery can re-traumatise victims and survivors of abuse. The IWF found that 96% of the reports processed in 2022 depicted exclusively girls. • Hash matching and URL detection can be useful and effective tools for combatting the circulation of CSAM for user-to-user and search services, respectively. • Hash matching involves analysing images and videos communicated publicly on the service and comparing a digital fingerprint of that content to digital fingerprints of previously identified CSAM. URL detection enables providers to ensure that users do not encounter, in or via search results, search content present at or sourced from URLs on a list of URLs previously identified as hosting CSAM.		

Case study 11: Gender-sensitive recommender system algorithms • A growing community of misogynistic influencers (sometimes referred to as 'misogyny influencers') can have considerable influence over the propagation of misogynistic content. Online misogyny can glorify, justify, and create tolerance for sexual violence. • Evidence shows that recommender systems reward influencers creating misogynistic content with greater reach, particularly to boys and young men. This happens because algorithms are optimised for high engagement, which over time can incentivise the production and exposure to polarising and harmful content. Furthermore, recommender systems can also disproportionately show other kinds of harmful content to girls and young women, such as content promoting eating disorders and self-harm. Such recommendations can be a result of embedded data bias in data service providers collect about users. Gendered bias and gendered disinformation can also be shared via GenAI chatbots and voice assistants which can replicate biased algorithms and training data. • Gender-sensitive approaches can reduce the spread of online misogyny, including abusive and hateful content. Training content recommendation algorithms to be gender-sensitive could include: • > Auditing and evaluating recommender algorithms and other AI systems to assess		Refuge recommends Add to recommended actions after *Training content recommendation algorithms to be gender-sensitive could include:* *> employ VAWG experts and survivors, including domestic abuse specialists, to provide up-to-date information and safety advice on the harms being experienced by women and girls and to advise on audit, evaluation and retraining of algorithms to create gender-sensitive systems.*

whether they promote online misogyny, as well as evaluating gender bias in recommendations. • > Retraining algorithms either after revising existing datasets by, for example, applying pre-processing bias-mitigation strategies, or training the algorithm (or 'classifier') on a new training dataset put together by a diverse group that includes humans with a high level of sensitivity and training on gender-based harms, including intersectional aspects. This may also include giving human annotators sufficient time to evaluate content carefully, especially when the evaluation needs to consider the context within which the content was posted.		
Case study 12: **Gender-sensitive search services** • Some websites and forums are dedicated to allowing users to create non-consensual intimate content, including nudification apps sexualising deepfakes.199 Providers of general search services can reduce access to these websites, forums, and applications to help protect individuals and society from illegal and harmful non-consensual material. This can include: > Delisting: action that results in the content no longer appearing in search results. > Deprioritising: ensuring that a particular piece of content is deprioritised in the overall ranking of search results and is therefore less discoverable to users.		Refuge is in favour of delisting and de-prioritising as way of preventing harm. We have commented above on the importance of service providers employing safe practices in handling requests for removal. There is an urgency for swift action to ensure that content sharing is minimised and for the people reporting to be informed about the action taken

> Reporting: making it easier for people to request removal of non-consensual content from search results		
Case study 13: Nudging to deter uploading of harmful content Introducing deliberate friction using nudges aimed at potential perpetrators can discourage uploading harmful behaviour or content without blocking it. A popular example implemented by a range of services is 'preliminary flagging.' When a user attempts to post harmful content, a moderation algorithm classifies the content as harmful or violating a service's community guidelines. Deterrence messaging can also be deployed where potentially harmful behaviour (rather than content) – such as repeatedly messaging a user they haven't engaged with before without a response – is detected. For example, a dating service could use deterrence messaging to reduce harmful interactions. Within a conversation, if harmful content is detected in a message, the sender could be prompted with a warning to reconsider the language used and be given the option to edit the message before sending. Deterrence messaging can be used in tandem with supportive messaging (for more information on supportive messaging, see Chapter 5). For instance, if a user who is prompted with a warning message opts to send the content anyway, then the user who receives the message could be prompted with a supportive message with information about how to report		Refuge believes that deterrence messaging is useful in identifying actors engaged in online VAWG. Tech companies should monitor and review the content being flagged on their site, the actions of the user receiving them, the impact of the support messages being sent to recipients. Refuge recommends adding Deterrence messaging will always be limited in deterring highly motivated offenders, and in identifying subtle forms of abuse (such as coercive control) With reference to deterrence messaging as follows: We recommend that the following is added: But it is still important to show that alleged perpetrators have had the potential harm clearly stated to them, as it could help prove intent for prosecution and restraint purposes.

and block other users on the service. Feedback loops between user reporting and deterrence messaging can improve both the efficacy of content moderation and deterrence messaging While such nudges show promising effects in improving online pro-social behaviour, existing studies emphasise the importance of getting the messaging right. Particular care needs to be taken with deterrence messages shown to children so that the messages do not lead children to be ashamed and avoid discussing negative online experiences with their parents or guardians. Deterrence messaging will always be limited in deterring highly motivated offenders, and in identifying subtle forms of abuse (such as coercive control or misgendering). There is also the risk of the impact of such nudges reducing over time as users become accustomed to the prompts and automatically click-through them. An intervention with a learning component or rotating several nudge messages could help to maintain effectiveness.		
Case study 14: Preventing image-based sexual abuse on adult services Providers of adult content services face higher risks of hosting non-consensual intimate content because their sites allow sexual content. Where a service provider identifies it is at risk of hosting nonconsensual intimate content, there is a variety of measures it can implement to mitigate this risk, such as:		See point made against 4.40(b) about issues with User ID verification With reference to the following: This can involve cross-industry initiatives such as StopNCII.org, which allow survivors and victims to generate hashes from their intimate images. These hashes are shared across participating service

Persuasion: • Consent nudging: Implementing a prompt asking the user if they have asked for consent from all parties depicted within the content (if the algorithm detects more than one person). A study by the Cyber Civil Rights Initiative found that 66% of perpetrators listed "if I had taken more time to think about what I was doing" as a reason that would have stopped them from posting nonconsensual images. • Uploader verification. Users must verify their identity to upload content, providing a full legal name, date of birth, a piece of matching government-issued photo ID, and a live face scan check. This can also include removing historic videos from unverified accounts. • Deterrence messaging: Warning messages about the illegality and consequences of intimate image abuse. **Removal:** • Hash matching. An automated system cross-references uploaded content against a database of hashes for previously reported non-consensual intimate images, with matches removed and prevented from being shared. This can involve cross-industry initiatives such as StopNCII.org, which allow survivors and victims to generate hashes from their intimate images. These hashes are shared across participating service providers to detect and prevent the circulation of these images. • **Consent verification**. Users must certify that all individuals depicted in uploaded content have consented		providers to detect and prevent the circulation of these images. and footnote 274 Ofcom has not assessed this particular tool for accuracy, effectiveness, and freedom from bias. Refuge recommends that Ofcom does assess the tool and that the assessment considers the impact of NCIs on survivors of domestic abuse.

to appear and must provide identity verification for those depicted. Service provides can use facial recognition and nudity detection to block uploads of content if the uploader cannot provide proof of consent. This approach could also allow depicted users to withdraw consent, especially in content involving nudity. Providers can layer different techniques to prevent intimate image abuse. Service providers may also refer to the Image-Based Sexual Abuse Principles on preventative approaches for the development of industry best practices		
Case Study 15: Automated detection of misogynoir content and results Digital misogynoir refers to online hate and dehumanising language experienced by Black women online, particularly on social media services. Although some of this content is likely to be illegal hate speech, existing automated hate speech detection tools are ineffective at detecting it due to a lack of sensitivity to context. This issue is particularly likely when such algorithms are trained on datasets which are tagged as just racist speech or misogynistic speech, therefore missing the intersections between the two. The effectiveness of these tools can be strongly influenced by the identity of annotators labelling hate speech, as well as other decision-makers within service providers. Researchers at Glitch, an online safety charity, and the Open University have been developing methods to better		

detect misogynoir, including through training on datasets of self-reported misogynoir. For such techniques to be successful, they need to recognise that safety measures which treat different kinds of abuse (such as racism and misogyny) in isolation will fail to account for intersectional hate. Digital misogynoir can also occur in search services. For example, researcher Safiya Noble found that searches for the term "Black girls" were more likely to result in pornographic results and sexually explicit terms than searches for "white girls". Likewise, searches for variations of "Black women" led to racist and sexist suggestions in autocomplete. To address these harms, search services could monitor their systems for embedded bias		
SUPPORTING WOMEN AND GIRLS		
Action 7: Give users better control over their experience		
Case study 15: Disabling comments • Pile-ons are a common type of coordinated harassment where a large group of users target either an individual, or a much smaller group of users. Perpetrators of pile-ons intend to shame and silence the individuals they target, particularly women and girls in public life such as politicians, journalists, activists, and athletes. • Comment features can be used by perpetrators in pile-ons to respond directly and publicly to other users' posts with threatening or abusive messages. • Enabling users to disable comments on their own posts (before or after posting) means that women and girls can		

respond to this harm when it occurs, or when they judge that it is at risk of occurring. > A user could disable comments on a post during a pile-on to hide any existing comments and prevent any further comments. > A user who is concerned about experiencing a pile-on could disable comments before posting to prevent this harm. • Women and girls can disable comments on their posts without having to restrict the visibility of their posts		
Case study 16: Supportive information • Online misogyny circulates through a wide range of content online, from posts on dedicated forums trivialising sexual assault to viral videos on social media sites that glorify domestic abuse. This type of content can cause harm and evidence shows it can lead to women and girls withdrawing from online participation. • While a small number of users deliberately search for this content, many encounter it unintentionally through content recommender systems, including boys. • Signposting to supportive information that is clear and accessible can increase users' awareness of the user control tools available to them and encourage users to consider their safety online. > A user who encounters a misogynistic video online and reports the content could receive supportive information. This information could set out steps they can take to further restrict content, accompanied by an explanation of the kind of content they may be restricting.		We recommend that this case study emphasises the importance of provision of easy-to-understand and accessible information (i.e. it can be easily be found on the providers platform and can be understood by all users). We recommend that tech companies are encouraged to refer to third party support services and provide accessible guides to specialist support themselves. When signposting tech companies should be encouraged to provide cautions for users about how to access information and support safely i.e. not from a compromised account or device where an abuser could see what the survivor is doing. See refugetechsafety.org site for example caution wording. We also recommend that tech companies provide clear steps about how to report this content and the

• Informed choices about content restriction can prevent users being exposed to further harm.		timeframe in which action will be taken and when and how they will be kept informed.
Case study 17: Mass blocking When online gender-based harms occur, survivors and victims should be able to limit their interactions with and exposure to perpetrators. The ability to block other users provides women and girls with greater control over who can follow their accounts, who can interact with their posts, and who can directly message them. A social media platform could provide users with more extensive blocking options. For example, the option to block not only another user's account but also any other accounts the user might have, as well as any new accounts the user may create in future. The platform could also allow users to block any current or future accounts connected to a particular phone number or email address. A social media platform could also offer users more automated blocking options. For instance, if a user sees a post that is offensive or disturbing to them, they could be given the option to block not only the post's author but all users who have reposted. Providing users with additional blocking options reduces the burden of safety work on survivors and victims, and creates friction for perpetrators attempting to evade blocks by creating new accounts. Providing women and girls with greater control over which users they can restrict interactions with enables them to		

reduce their risk of experiencing online gender-based harm		
Case study 18: Content filtering Filters give users greater control over their experience online and prevent them from encountering unwanted content, including the use of words and phrases which amount to online gender-based harm. Content filters can provide users with control in a variety of ways, such as by allowing them to reduce the amount of violent or sensitive content they are shown. A social networking platform could provide content filtering tools that allow users to identify topics they do not want to engage with by flagging specific tags, keywords, and phrases they do not want to see. Content filters are not usually case sensitive, and keyword filters should also work on any terms which include the keyword. It takes time for users to set up filters, but it allows them to personalise the content they see. The tool empowers women and girls to shape their online experiences and avoid content which contains words and topics likely to be offensive, disturbing, or upsetting to them.		
Action 8: Enable users who experience online gender-based harms to make report		
Case study 19: Affected persons • Intimate image abuse can include the non-consensual sharing of both images created consensually and images created non-consensually, such as deepfakes. • User reporting is an important mitigation against intimate image abuse, especially for services on which		We recommend adding as follows: This reduces the risk of re-traumatisation. But in some instances, women and girls are terrified to let anyone else know. So, it should be paramount that

users can encounter nudity and sexually explicit content. • Allowing affected persons to make complaints makes it easier for women and girls who experience intimate image abuse to report it. > A survivor and victim can report intimate image abuse depicting them as an affected person, even if they are not a user of the service it has been uploaded to. This means that intimate image abuse can be reported without the survivor and victim having to create a user account to access the service's reporting system. > A survivor and victim of intimate image abuse can ask another individual to report images on their behalf. This reduces the risk of re-traumatisation. • Complaints processes reduce friction for survivors and victims in reporting intimate image abuse		tech companies allow people to report without having an account. The reporting process should use blame free language and be mindful that the reporter may be experiencing different forms of domestic abuse simultaneously and provide safe and responsible signposts to appropriate resources.
Case study 20: Reporting options • Harassment is an offence involving a course of conduct, which includes causing another person alarm or distress. The perpetration of harassment is often highly personal and can involve patterns of behaviour aimed at isolating the survivor and victim. • This behaviour does not always occur through written and visual communication such as images, comments, and direct messages which can be easily recorded and reported. • Offering accessible reporting for all types of content and interaction supported on a service ensures that women and girls are always able to report harassment to providers.		

> A user on a gaming service can report harassment when it happens during in-game voice chat. This prevents perpetrators subverting the service's enforcement systems by using voice chat. > A user on a virtual reality service can report a perpetrator sexually harassing them and invading their physical space. • Reporting systems must be updated to cover any changes providers make to their services, including new possibilities for user interaction		
Case study 21: **Track and manage reports** Experiences of online gender-based harms are often complex and highly contextual and frequently involve multiple interactions or pieces of content. Reporting is time-consuming and can be re-traumatising for survivors and victims. Reporting systems that allow users to track and manage their reports can provide survivors and victims with greater agency and transparency over the process. The Web Foundation's Tech Policy Design Lab developed a prototype for a reporting dashboard that enables users to track their reports and see when reports are resolved. The dashboard could allow users to add additional context to their reports and collect and archive evidence of harmful content. Providers could also give users the option to invite a trusted contact to support with the reporting process. Providing users with greater choice over the reporting process allows women and girls to tailor the reporting process to their experiences and		Refuge recommends that the ability for survivors to download a copy of their reports be added. This links to the points set out above, and in response to question one on the need for tech companies to better facilitated survivors gathering evidence of the abuse being perpetrated against them, which could provide enormous benefit to survivors navigating different system and agencies where she is required to prove that abuse is perpetrated against her. This data could also be used as evidence in some criminal and civil cases.

preferences. This can help overcome challenges in the reporting of online gender-based harms and build trust between providers and survivors and victims.		
Case study 22: Trusted flaggers Reporting online gender-based harms and engaging with providers can be a challenging process for individual survivors and victims. Trusted flagger programmes can assist with this process by building relationships between providers and organisations with expertise in harms such as online domestic abuse and intimate image abuse. A 'Violence Against Women and Girls Code of Practice' developed for industry by a civil society coalition recommends that providers set out clear criteria for what content trusted flagger organisations can report and provide a specific route for escalation if providers do not respond to trusted flagger reports. The coalition also emphasise that trusted flagger organisations should be provided with the necessary resource and support when carrying out additional work to make a service safer. Developing partnerships between providers and organisations with expertise in gender-based harms gives survivors and victims additional support and advocacy. These partnerships can also be used to alert providers to emerging forms of harm		Refuge's tech team has developed relationships with many technology companies and has been recognised as a Trusted Partner by several major social media platforms. Trusted Partner programmes enable charities and researchers to communicate directly with safety teams at social media companies, to report abusive content and, in theory, to receive a more rapid response from the platform. (excerpt from Refuge's Marked as Unsafe report). The Refuge Tech Abuse team gave an example of this arrangement working well: A survivor was out of their Snapchat account. The perpetrator had access to the survivor's account, and, in this case, they were sharing intimate images of the survivor. The survivor was unable to stop the perpetrator as they no longer had access to the account- this had been happening for up to a year. Through the trusted flagger process the Refuge team were able to highlight the account to the platform, which then removed the content and, in this case, stopped the perpetrator from having access to the account. Some of the images may be used in a future police investigation. Having the direct path to the platform was integral to action being taken, as those images had been up for some time without the survivor being able to prevent them from being

		shared. If a survivor no longer has access to an account, it can be difficult for them to remove images, so having the trusted flagger relationship to advocate for the survivor is very important.
Case study 23: Reporting off-service behaviour Online gender-based harms are not always restricted to a single interaction or piece of content. Harms such as stalking are part of a wider pattern of behaviour, both online and offline. Reporting systems which allow survivors and victims' to flag gender-based harm that has happened offline or on another service enable providers to reduce the risk of harm occurring on the service. A livestreaming service could introduce a policy that enables users to report harmful off-service behaviour. This would allow users of the service who have experienced stalking offline to inform the provider. The provider could investigate and, if satisfied that sufficient verifiable evidence has been given, the provider may take enforcement action. Enforcement action could include blocking the perpetrator from interacting with the survivor and victim on the service to prevent any future abusive behaviour. Accounting for off-service instances of gender-based harms such as stalking helps tackle harmful patterns of behaviour and enables providers to take proactive action to prevent online gender-based harms occurring on their service		
Action 9: Take appropriate action when online gender-based harms occur		
Case study 24: Domestic abuse training		

• Domestic abuse is perpetrated in complex and highly personal ways, and online domestic abuse often replicates and extends the same dynamics as offline domestic abuse. • User reports of online domestic abuse are contextual and may focus on a pattern of behaviour rather than a single interaction or piece of content. • Training content and search moderation teams on domestic abuse enables moderators to better identify instances of this harm and respond to user reports. > Trained content and search moderation teams could take into account considerations such as how to respond to user reports appropriately without escalating offline risks to survivors and victims. > Providers, specifically content and search moderation teams, could develop their understanding of domestic abuse and how it occurs on their service through partnerships with organisations who have frontline experience and expertise. • Content and search moderators should receive adequate support and safeguarding from providers to undergo training and carry out this work		
Case study 25: Action on serial perpetrators A small number of users are often responsible for a large amount of online gender-based harm, particularly in cases of co-ordinated harassment. Evidence shows these users engage in repetitive and abusive behaviour which targets women, such as repeatedly posting the same sexually explicit content.		Refuge recommends that the following is added Providers should take into account the impact of the online behaviour on domestic abuse survivors; assess the risk and impact of reinstating the user; set up a process for informing the survivor or bystander reporter if they have reported the problem.

A social networking service with a Generative AI feature could implement a strike-based enforcement system, where a user receives a 'strike' against their account for misuse of the service. For instance, if a user attempts to generate harmful content through the Generative AI feature, they could receive a strike against their account. If a user receives multiple strikes, their access to the Generative AI feature could be removed for a given period of time. If a perpetrator continues to misuse the service, repeated strikes could lead to an account ban. Users should be informed when they receive a strike and what the consequences of a strike are. Providers should also include the ability for users to appeal strikes and related enforcement action. Applied effectively, strike-based enforcement systems can act as a form of deterrence to prevent a single act of abusive behaviour from becoming a pattern. It is worth noting that content moderation and tools assessing user behaviour are likely to involve processing of personal data. This includes moderation actions applied to a user's account (such as a strike, service restriction or ban). We encourage services to consult guidance from the Information Commissioner's Office.		