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

Consultation response form

Your response

Question	Your response
Question 1: Do you have any comments on our proposed approach to 'content and activity' which 'disproportionately affects women and girls'?	Confidential? – N We welcome Ofcom's publication of the draft Guidance. However, prior to turning to specific questions, we want to highlight the fundamental obstacle to making this Guidance transformative in relation to improving women's online safety - namely, that this Guidance is completely voluntary and based on the will of individual platforms and websites to incorporate this Guidance into their policy. It is non-enforceable against them and the voluntary nature is reinforced through the adopted language throughout the Guidance. For example, Consultation Document A in section 2.20 clearly states that 'service providers can use their discretion to determine which solutions will be most relevant to meet their illegal harm and protection of children duties and be most impactful for their users'. [1] Considering that a

[1] OFCOM, 2.20 Consultation Document A, p. 17

Question	Your response
	significant part of the implementation of the Guidance is left to platform providers' discretion and is therefore legally non-enforceable, it is unlikely that the Guidance will make a substantial difference to the behaviour of platform providers as they will be able to either ignore the Guidance completely or to implement measures that are least effective thereby diluting the effectiveness of OFCOM Guidance and the Online Safety Act 2023 further. To ensure the adoption of good practice by online platforms and the prevention of harm to women and girls, the Guidance must be transposed into an enforceable Code of Practice as soon as practicable. The fast-moving world of online safety makes the extension of legislation and regulation inevitable. Centre for Protecting Women Online (CPWO) is of the view that a Code of Practice would have provided a much more legally robust and enforceable basis for improving women's online safety. We respond to this Consultation with a hope that these recommendations will, in the future, evolve into a Code of Practice. CPWO is happy to lend our expertise and be involved in the future as this work develops.

Question	Your response
	We also have concerns with effective implementation of the Online Safety Act (OSA 2023). Ofcom's interpretation of the Act so far – for example, in relation to the harms it asks services to act on, its application of the principle of proportionality to the measures required of services – has been criticised by experts,[2] civil society organisations[3] and in Parliament[4]. For example, End Violence Against Women Coalition has published an open letter stating that Ofcom's regime is 'unlikely to fulfil the potential of the law to tackle online' violence against women.[5] Meanwhile, to provide evidence of the need for an enforceable Code of Practice, Ofcom must put in place a robust system of monitoring and evaluation to document the take up, and lapses in take up, of the voluntary Guidance by service providers. Further, we are also concerned about the balancing of women's human rights and other potentially conflicting rights (e.g. rights of other users, interested parties

[2] For analyses and critiques of Ofcom's approach to implementation by the Online Safety Act Network, see https://hidden-bayou-38064-8bb90f096618.herokuapp.com/analysis/types/consultation-responses/.

[3] End Violence Against Women Coalition, 'Ofcom blocking a safer internet for women and girls, VAWG experts warn' (EVAW, 23 February 2024) available at: https://www.endviolenceagainstwomen.org.uk/ofcom-blocking-a-safer-internet-for-women-vawg-experts-warn/

[4] HC 26 February 2025, vol 762, available at: https://hansard.parliament.uk/commons/2025-02-26/debates/00959567-3EC5-4AF6-94DC-A242D0EE1B0A/OnlineSafetyActImplementation

[5] End Violence Against Women Coalition, Open Letter to the Chief Executive of Ofcom, (EVAW, 23 February 2024) available at: https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/2024/02/VAWG-letter-to-OFCOM.pdf

Question	Your response
	and services) as expressed on pp. 48-49 of the Consultation Document.[6] It is not clear how this balancing will be carried out in specific cases, who will do the balancing, whether Ofcom will engage with external experts, how much weight will be given to conflicting rights claims, and so on. Our concern is that, without the involvement of women's human rights experts in this process, women's rights may be given less weight than non-absolute rights such as freedom of speech. This is a crucial issue given the recent backsliding in relation to fact checking moderation of content[7] in the name of promoting freedom of expression[8]. **Question 1: Do you have any comments on our proposed approach to 'content and activity' which 'disproportionately affects women and girls'?** Having reviewed the consultation documents, we have concerns in relation to the way Ofcom has categorised 'content and activity' which 'disproportionately affects women'. We believe that having only four categories (i.e. online misogyny, pile-ons

[6] OFCOM, Consultation Document: A Safer Life Online For Women and Girls, pp. 48-49 available at: https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-on-draft-guidance-a-safer-life-online-for-women-and-girls/main-docs/consultation-document-a-safer-life-online-for-women-and-girls.pdf?v=391803

[7] L. McMahon, Z. Kleinman and C. Subramanian, 'Facebook and Instagram get rid of fact checkers' (BBC News, 7 January 2025) available at: https://www.bbc.co.uk/news/articles/cly74mpy8klo

[8] J Kaplan, Chief Global Affairs Officer, More Speech and Fewer Mistakes (Meta, 7 January 2025) available at: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/

Question	Your response
	and online harassment, online domestic abuse and image based sexual abuse) is too restrictive and does not appropriately capture the full range of harmful behaviours that women experience online nor the online gender-based harms resulting from those behaviours. We believe that this will not enable platform providers to capture numerous other harms that disproportionately affect women.[9] For example, it is unclear if harms such as catfishing, digital sex trafficking, threats of violence and doxxing would be captured within any of these categories. Section 2.9 of Annex A of Draft Guidance states online misogyny is a wide range of content and behaviour online which engages in, normalises or encourages misogynistic attitudes and ideas.[10] This definition is ambiguous. Whilst the Annex A indicates forms of online misogyny are listed within the Illegal Harms Register of Risks, these harms are still not highlighted as being online misogyny. This ambiguity means it is unclear what will constitute online misogyny, meaning enforcement of

[9]K. Barker and O. Jurasz, Text-Based (Sexual) Abuse and Online Violence Against Women: Toward Law Reform in J. Bailey, A. Flynn, and N. Henry (eds), *The Emerald International Handbook of Technology Facilitated Violence and Abuse* (Emerald Publishing, 2021) pp. 247–264. https://doi.org/10.1108/978-1-83982-848-520211017

[10] Section 2.9, Annex A Draft Guidance (Ofcom, February 2025) available at: https://www.ofcom.org.uk/siteassets/resources/documents/consultations/category-1-10-weeks/consultation-on-draft-guidance-a-safer-life-online-for-women-and-girls/main-docs/annex-a-draft-guidance.pdf?v=391669

Question	Your response
	OFCOM Guidance will not be as effective as it could be. This is exacerbated by the fact that 'misogyny' and/or 'online misogyny' are not defined in any of the laws across the UK nor in international treaties binding the UK. 'Online misogyny is a form of gender-based cyberhate, directed against women because they are women'.[11] Aggression, bullying and online harassment can also be a form of online misogyny. Feminist theory describes misogyny as a range of activities from hostility towards women to physical, psychological and systemic violence against them.[12] It can include distributing content related to belittling of women, exclusion of women in society and promotion of patriarchy or male privilege.[13] Online misogyny utilises and promotes gender stereotypes and patriarchal norms. It affects women of all backgrounds, who participate online, including in situations when they express their views and where these opinions represent a feminist or otherwise not mainstream

[11] K. Barker, O Jurasz, Online Misogyny as a Hate Crime: A Challenge for Legal Regulation? (Taylor Francis, 2018) p.xiv available at: https://www.taylorfrancis.com/books/oa-mono/10.4324/9780429956805/online-misogyny-hate-crime-kim-barker-olga-jurasz?refId=7f7c162f-836a-4149-8e28-903b9b7dbc69&context=ubxl

[12] M. Duggan, H. Mason-Bish, *A feminist theoretical exploration of misogyny and hate crime*, (Routledge, 2021) pp. 2–6 available at: https://core.ac.uk/download/pdf/337607827.pdf

[13] L. Code, Encyclopaedia of Feminist Theories (Routledge, 2001)

Question	Your response
	viewpoint.[14] Further, there are specific forms of misogyny experienced by Black women, which are coupled with racism.[15] As a result, it is necessary to ensure that misogyny is effectively addressed by the Draft Guidance. Given this ambiguity, platforms might be able to argue that content is not deemed as misogynistic and does not 'violate community standards' which is further amplified by Meta's decision to limit fact-checking.[16] For example, content that is sexist, which may be dismissing or belittling a woman online, for instance, due to her occupation in law enforcement may not necessarily fall within the misogynistic category. Meta explicitly states their policies allow room for sex and/or gender exclusive language when discussing topics such as law enforcement[17]. Further, recent research demonstrates that platforms' detection algorithms are often ineffective at detecting misogynoir.[18] This is because

[14] K. Barker, O Jurasz, Online Misogyny as a Hate Crime: A Challenge for Legal Regulation? (Taylor Francis, 2018) p.25 available at: https://www.taylorfrancis.com/books/oa-mono/10.4324/9780429956805/online-misogyny-hate-crime-kim-barker-olga-jurasz?refId=7f7c162f-836a-4149-8e28-903b9b7dbc69&context=ubxl

[15] J. Kwarteng et al, Misogynoir: challenges in detecting intersectional hate (2022) Social Network Analysis and Mining, available at: https://oro.open.ac.uk/85874/8/s13278-022-00993-7.pdf

[16] M. Verveer and K Baekgaard, 'Meta's Move to Limit Fact-Checking Endangers Women-and Democracy' (*Lawfare,* 6 March 2025) Available at: https://www.lawfaremedia.org/article/meta-s-move-to-limit-fact-checking-endangers-women-and-democracy

[17] 'Hateful Conduct' (*Meta Transparency Center*, April 2025) available at: https://transparency.meta.com/en-gb/policies/community-standards/hateful-conduct/

[18] J. Kwarteng et al, Misogynoir: challenges in detecting intersectional hate (2022) Social Network Analysis and Mining, available at: https://oro.open.ac.uk/85874/8/s13278-022-00993-7.pdf

Question	Your response
	these systems rely too heavily on explicit language to determine harm, whilst missing content that requires knowledge of specicific context and understanding of misogynoir.[19] Understandably, as mentioned in section 1.20 of Annex A, harms are evolving rapidly. However, having clear and comprehensive definitions of harms constituting online misogyny with clear examples of such harms, which could be adapted as harms evolve, reduces confusion or technicalities that will be faced by providers. For example, some crimes, such as stalking, can be made up of several online communications that on their own may not be deemed misogynistic. As a result, there is a risk that these would not be deemed to fall within the online misogyny category and may not be addressed by platforms. *Pile-ons and online harassment* While the depiction of pile-ons and harassment are accurate within Annex A, online platforms may contend that numerous messages or comments lack a direct connection, making it challenging to associate them to coordinated behaviour.

[19] J. Kwarteng et al, Misogynoir: challenges in detecting intersectional hate (2022) Social Network Analysis and Mining, available at: https://oro.open.ac.uk/85874/8/s13278-022-00993-7.pdf

Question	Your response
	Annex A (2.23) includes a well-articulated explanation of the range of behaviours that constitute pile-ons and harassment as well as which demographics are more at risk of being subjected to said behaviours. Although women in the public eye face heightened risks of pile-ons, there appears to be less attention given to those outside of the public sphere who nonetheless experience pile-ons. As mentioned in Annex A (2.23), women are at risk of pile-ons simply by having an online presence. While it is crucial to prioritise protections for those most affected, measures should be reflective of the need to safeguard *all* women who experience pile-ons regardless of their profession. This is inclusive of gendered mis/disinformation. Again, this lack of clarity will likely result in challenges in enforcing any Guidance published by OFCOM. *Online Domestic Abuse* Annex A outlines a definition of online domestic abuse, but it exclusively focuses on abuse involving a victim and their current or former partner. This contrasts with other definitions that also encompass perpetrators such as family members[20]. The list of

[20] 'What Is Domestic Abuse?' (*Women's Aid*, 3 October 2024) Available at: https://www.womens-aid.org.uk/information-support/what-is-domestic-abuse/;

Question	Your response
	harms provided is accurate; however, many of these are not unique to online domestic abuse. For example, some pile-ons and doxxing is domestic abuse related, and other examples are likely to be stranger perpetrators. Agencies working with survivors, such as Refuge, have identified a link between the ex-partner "recruiting" others to harm and abuse the survivor online. These can be family members, friends or associates, or people they meet online. It is noted in section 2.29 of Annex A that a picture of a front door can seem innocuous but may trigger the feeling of being unsafe to domestic abuse survivors. While this is accurate, it is also the case for women who are not subjected to domestic abuse but are victims of online violence committed by unknown perpetrators. Broadening the categorisation of some harms will allow for wider protections for women online. *Image-Based Sexual Abuse* In stark contrast to other categories, the harms outlined under image-based sexual abuse are both accurate and presented in a clear, coherent manner. How-

Question	Your response
	ever, it may be valuable to include sextortion and pressurised sexting, which involves non-consensual solicitation of intimate images, as this issue is often overlooked in discussions surrounding image-based abuse.[21] Annex A (2.33) references intimate images in the context of culture and religion but does not provide further information. Expansion of this aspect is essential to support effective enforcement of further Guidance and promotion of culturally sensitive understanding of what constitutes 'intimate'. For example, threatening to share or sharing images of a Muslim woman without a hijab will cause significant stress and anxiety and, again, these scenarios are overlooked[22]. Ofcom should develop further understanding of what the term 'intimate' means. This is necessary to ensure platforms do not exclude content that can be just as harmful to women due to a focus on the term 'sexual'.[23] For example, women form conservative religious backgrounds could be harmed as a result of an image depicting them hugging a man or sitting on a bed with a man. Many technology platforms

[21] What is Sextortion, (Internet Matters) available at https://www.internetmatters.org/resources/what-is-sextortion/, Sextortion, (The Metropolan Police) available at: https://www.met.police.uk/advice/advice-and-information/online-safety/online-safety/sextortion/

[22] A Waheed, 'For Muslim Women, Images of Us without Our Hijabs Can Be as Damaging as Nude Photos' (*Glamour UK*, 24 March 2025) Available at: https://www.glamourmagazine.co.uk/article/muslim-women-image-based-abuse-hijab

[23] H Hussain, 'Chayn is Building a Cultural Map of What 'Intimate' Means Around the World' (Chayn, 5th March 2025) available at: https://blog.chayn.co/chayn-is-building-a-cultural-map-of-what-intimate-means-around-the-world-400436f39eb4

Question	Your response
	would not consider these situations harmful enough to take down, but they are harmful and can even be lethal.[24] Ofcom Guidance should emphasise that cultural differences and attitudes towards women's bodies will have an impact on what images are considered intimate. Although it is understandable that any guidance will not encompass every potential harm, some refinements could help address abuses that may currently be overlooked or unconsidered.
Question 2: Do you have any comments on the nine proposed actions? Please provide evidence to support your answer.	Confidential? – N The nine proposed actions are split into a series of measures categorised under 'foundational' and 'good practice', with the latter section setting out examples of technologies that can be used by tech platforms to protect women and girls. For example, the use of hashing technology to take down intimate images posted without consent on their platform. However, considering that both 'foundational' and 'good practice' categories are completely voluntary, this distinction seems to be irrelevant and is highly unlikely to make any

[24] H Hussain, 'Chayn is Building a Cultural Map of What 'Intimate' Means Around the World' (Chayn, 5th March 2025) available at: https://blog.chayn.co/chayn-is-building-a-cultural-map-of-what-intimate-means-around-the-world-400436f39eb4

Question	Your response
	difference to behaviour of platform providers. As a result, the purpose behind the distinction between foundational and good practice steps is unclear. Considering that foundational actions are not legally enforceable, it is unclear what this distinction is trying to achieve in practice. It would be recommended to ensure that the 'foundational' actions are related to the legally enforceable Codes of Practice and the Illegal Harms Register of Risks. If foundational actions would be compulsory and legally enforceable, whereas the 'good practice' actions would be voluntary, this distinction would make sense and would be coherent. However, as both foundational and good practice actions are completely voluntary, the 'good practice' actions appear to water down further responsibilities of platforms.
Question 3: Do you have any comments about the effectiveness, applicability or risks of the good practice steps or associated case studies we have highlighted in Chapter 3, 4 and 5? Are there any additional examples of good practices we	Confidential? – N <u>Chapter 3, Action 3: Be transparent about women and girls' online safety</u> Transparency requires the ability to systematically document the relationship between online safety requirements and technology features of online services in ways that allow users to analyse this doc-

Question	Your response
should consider? Please provide evidence to support your comment.	umentation to identify how factors of concern to them are addressed by the system. Prior research in security requirements provides some useful starting points for documenting abusability and analysing how the interactions between system features can lead to harms[25]. Online services should also help users who report online harms to be kept informed on the progress of any investigation into their reports and the action that was taken to address both the specific incident and to mitigate future risks to other users. Public registers of how online harms were perpetrated using different technology features of online services could also help software product teams improve their systems to minimise the risks of harm more effectively. Such registers could be based on similar approaches used in the cyber security domain, such as MITRE's Common Attack Pattern Enumeration and Classification (https://capec.mitre.org) and the ATT&CK taxonomy (https://attack.mitre.org). Chapter 4, Action 4: Conduct abusability evaluation and product testing

[25] C. Haley, R. Laney, J Moffett, J. and B, Nuseibeh, Security requirements engineering: framework for representation and analysis (2008) IEEE Transactions on Software Engineering, 34(1), pp.133-153.

Question	Your response
	Further, in relation to Chapter 4, Action 4, which deals with conduct abusability evaluation and product testing, we are concerned that there is no current standard for abusability evaluations for mitigating online harms. However, abusability evaluations seem analogous to vulnerability assessments that are conducted on systems to mitigate cyber security risks. Therefore, techniques like misuse case modelling[26] and attack trees[27] could provide a basis for developing methods for abusability evaluations. A key difference is that in cyber security vulnerability assessments, the analyst is trying to determine how the attacker can achieve their goal of compromising different security objectives of the system – e.g., breaching confidentiality, integrity or availability of assets. In the case of abusability evaluations, it would be necessary to establish the goals and motivation of the potential perpetrators of online harms covered by the Online Safety Act 2023, as a basis for exploring the potential mechanisms through which a technology features of an online service could be abused to achieve these goals. The complexity of these evaluations will be compounded by how interacting features

[26] G. Sindre, and A.L Opdahl, Eliciting security requirements with misuse cases (2005) Requirements engineering, 10, pp.34-44
[27] B. Schneier, Attack trees (1999) Dr. Dobb's journal. 24(12):21-9.

Question	Your response
	of multiple online services lead to potential abuses, which requires a systems approach. It is also relevant to consider how risk assessments are carried out in the context of cyber security and privacy impact assessments. Risk should be considered the product of likelihood and impact of technology being abused to perpetrate online harms. Here, the likelihood is determined by the motivation of a perpetrator to perform a harmful act and the extent to which the online service enables this. Empirical measures of likelihood and impact could be gathered from reporting of online harms (e.g., to authorities, or platforms like the Cyber Helpline), but analysis of this data will need to consider the contextual factors associated with incidents to ensure risk is assessed accurately. Further work is needed to determine how generic risk management methods of avoidance, reduction, transfer and acceptance would be mapped to the domain on online safety. Relevant research questions to investigate include the scope and presentation of information to end users that would enable them to make informed decisions on accepting certain risks. This relates to the expectation in the Guidance of online service providers being transparent about women and girls' online safety.

Question	Your response
	Chapter 5. Action 9: Take appropriate action when online gender-based harms occur The 'Good Practice' steps in this section of the Draft Guidance lists numerous actions that providers can take, including *fact-checking and labelling of content* to address gendered misinformation. In relation to the labelling AI-generated content, our research shows that these interventions need to be implemented with care because labels will be believed by users irrespective of whether the content is real or AI-generated.[28] Therefore, mislabelling of content could lead to the unintended consequence of users mistrusting valid content.[29] Further, our research indicates that the presence of labels for AI-generated content does not dramatically alter user behaviour when it relates to sharing of the content or commenting on the relevant content. As a result, whilst labels are a valuable tool in raising awareness of AI-

[28] D. Gamage, D. Sewwandi, M. Zhang, M, A.K. Bandara, Labeling Synthetic Content: User Perceptions of Label Designs for AI-Generated Content on Social Media, (2025) (To appear) in Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI2025), Japan, pre-print: https://osf.io/preprints/psyarxiv/p5t3v_v2 p.11

[29] D. Gamage, D. Sewwandi, M. Zhang, M, A.K. Bandara, Labeling Synthetic Content: User Perceptions of Label Designs for AI-Generated Content on Social Media, (2025) (To appear) in Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI2025), Japan, pre-print: https://osf.io/preprints/psyarxiv/p5t3v_v2 p.11

Question	Your response
	generated content, they are not an effective solution for mitigating the risks posed by deepfakes.[30] We have substantial concerns about the effectiveness of the Guidance as well as of good practice steps contained therein. Considering that the good practice steps are completely voluntary, it is highly unlikely that they will be effective. The voluntary nature of the Guidance demonstrates a lack of ambition to effectively tackle the scale and breadth of harms that women and girls face online and, worryingly, deprioritisation of their impact on women and girls as well as broader society. As noted above, the 'good practice' areas appear to dilute the importance of some actions, because they are not deemed to be 'foundational'. It is unclear why some actions such as the use of hashing technology to take down image-based sexual abuse is not considered to be a foundational action for platforms to implement. We note that application of hashing technologies is one of the key requirements in the Ofcom's Illegal Harms Codes of Practice for user-to-user services in relation to

[30] D. Gamage, D. Sewwandi, M. Zhang, M, A.K. Bandara, Labeling Synthetic Content: User Perceptions of Label Designs for AI-Generated Content on Social Media, (2025) (To appear) in Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI2025), Japan, pre-print: https://osf.io/preprints/psyarxiv/p5t3v_v2 p.11

Question	Your response
	child sexual abuse material (CSAM).[31] We believe that similar mandatory measures need to be recommended for tackling image-based abuse material directed at women. Further, section 4.31 of Annex A provides that service providers can use persuasion (supportive or deterrence messaging), removal (preventing uploads or taking down content), or reduction (limiting circulation or reducing visibility) to reduce the circulation of harmful content or messages. At the same time, section 4.32 of Annex A clarifies that it is up to the services to decide which methods will be most appropriate in each case. However, leaving the selection of methods solely to service providers may pose significant risks. As highlighted in sections 4.43 and Case Study 15, relying only on one or a few approaches can lead to ineffective outcomes due to issues such as insufficient contextual understanding[32], failure to account for multiple marginalised identities[33] and inherent algorithmic biases.[34] Conse-

[31] Ofcom, Illigal Content Codes of Practice for user-to-user services, (OFCOM, 2025) available at: https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/information-for-industry/illegal-harms/illegal-content-codes-of-practice-for-user-to-user-services-24-feb.pdf?v=391889

[32] L. Gao, and R. Huang, Detecting online hate speech using context aware models (2017) *arXiv preprint arXiv:1710.07395*

[33] J. Kwarteng, S.C Perfumi, T. Farrell, A. Third, & M. Fernandez, Misogynoir: challenges in detecting intersectional hate (2022) *Social Network Analysis and Mining*, *12*(1), 166

[34] W. Yin, and A. Zubiaga, Towards generalisable hate speech detection: a review on obstacles and solutions. (2021) *PeerJ Computer Science*, *7*, e598

Question	Your response
	quently, service providers should be required to rigorously test and compare multiple automated techniques for both effectiveness and fairness before deciding which methods to implement. Moreover, providers should document and justify their choice of approaches. This will not only ensure that decisions are data-driven and transparent but also that they genuinely address the complex landscape of online gender-based harms and, therefore, ensure accountability for processes and outcomes. Additionally, as noted in 4.43b, it is essential that these automated systems are subject to periodic reviews and updated assessments. Given that the nature of hate speech and abusive content evolves over time, an approach that is effective today may become outdated tomorrow. Regular reporting and reassessment can help identify when current methods no longer meet efficacy or fairness benchmarks and when emerging techniques may offer improved performance. In summary, there should be a framework mandating systematic testing, justification, and continuous evaluation of automated detection approaches. This approach will help ensure that the measures

Question	Your response
	used are both effective in reducing harmful content and accountable to users and regulatory bodies.
Question 4: Do you have any feedback on our approach to encouraging providers to follow this guidance, including our proposal to publishing an assessment of how providers are addressing women and girls' safety? Do you have any examples or suggestions of other ways we could encourage providers to take up the 'good practice' recommendations?	Confidential? – N As previously mentioned, the effectiveness of this Guidance is considerably undermined due to its voluntary nature. Whilst the Guidance states in various places that Ofcom will not hesitate to use its 'robust enforcement powers' the Guidance lacks clarity on how and when these will be applied. This feels contradictory, considering that complying with both 'foundational' and 'good practice' actions as set out in the Guidance is discretionary for each service provider and hence, no enforcement powers would follow for non-compliance. Further, as the Guidance may not apply to smaller services or platforms with a minimal UK user base, this would leave multiple harmful online spaces unmonitored.[35] Unless the Guidance becomes a Code of Practice and platforms face significant financial penalties for failure to comply, Ofcom is unlikely to succeed at encouraging providers to take up this Guidance. To make any difference at all, it will be crucial for Ofcom to monitor and evaluate uptake of this Guidance.

[35] End Violence Against Women Coalition, Open Letter to the Chief Executive of Ofcom, (EVAW, 23 February 2024) https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/2024/02/VAWG-letter-to-OFCOM.pdf

Question	Your response
	Whilst the Guidance includes transparency measures and publishing an assessment of how providers are addressing women and girls' safety, this is likely to have limited effect in practice. It is questionable whether platforms will reveal all of the information, especially if it shows the ineffectiveness of their safety measures. Similar transparency requirements regarding content moderation can be found in the EU's Digital Services Act.[36] However, it is already becoming apparent that the transparency requirements have not had the desired effect, with platforms like TikTok, Google, and Instagram failing to provide data by Member State.[37] Significant disparities can exist in the granularity, consistency, and standardization of disclosures across platforms[38], with platforms being able to drown the reports with data, whilst revealing very little about how their safety measures or content moderation works in practice.
Question 5: Do you have any comments on our impact assessment, rights assessment, or equality impact assess-	Confidential? – N It is important that to mitigate and manage the risks of harm service providers address differential experiences of users with

[36] Article 15, Regulation 2022/2065 of the Eureopean Parliament and Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) available at: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

[37] A. Zornetta, Is the Digital Services Act Truly A Transparency Machine? (Tech Policy Press, 11 July 2024) available at: https://www.techpolicy.press/is-the-digital-services-act-truly-a-transparency-machine/

[38] A. Zornetta, Is the Digital Services Act Truly A Transparency Machine? (Tech Policy Press, 11 July 2024) available at: https://www.techpolicy.press/is-the-digital-services-act-truly-a-transparency-machine/

Question	Your response
ment? Please provide any information or evidence in support of your views.	specific characteristics who are particularly impacted by 'harm'. Providers need to consult with diverse groups from marginalised backgrounds, including BAME women. This is because survivors from marginalised groups, can actively shape AI systems through their insights. For example, recent research highlights that BAME women are more effective at detecting and annotating misogynoir (anti-Black racist misogyny) compared to other groups (e.g., white males) due to their lived experiences[39].
Question 6: Do you agree that our draft Guidance is likely to have positive effects on opportunities to use Welsh and treating Welsh no less favourably than English? If you disagree, please explain why, including how you consider the draft Guidance could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.	Confidential? – N

Please complete this form in full and return to OS-Section54@ofcom.org.uk..

[39] Kwarteng, S.C Perfumi, T. Farrell, A. Third, & M. Fernandez, Misogynoir: challenges in detecting intersectional hate (2022) Social Network Analysis and Mining, 12(1), 166 available at: https://dl.acm.org/doi/pdf/10.1145/3578503.3583612