Ofcom: New priority offences - serious self-harm and cyberflashing

Centre for Protecting Women Online Response to Ofcom Consultation: New Priority Offences – Serious Self-Harm and Cyberflashing

About the Centre for Protecting Women Online

The Centre for Protecting Women Online is focused on understanding and addressing challenges to women's safety online, working collaboratively across sectors to inform research, organisational practice and government decision making.

The Centre's aim is to reduce online harms suffered by women and girls, promote pro-social behaviours online and help build better and more responsible tech and software.

The Centre is led by Prof [redacted], Professor of Law and work of the Centre is delivered five interwoven work streams: Law and Policy, Human Behaviour, The Future of Responsible Tech, Ethical and Responsible Tech/AI and Policing.

The Centre is funded by a £7.7 million grant from Research England.

Consultation response

Summary of overarching themes

Implications of AI: How AI is changing, promoting and encouraging harms

Continuum of Abuse and Relation to Wider Forms of VAWG: How cyberflashing, self harm and suicide can be part of a broader pattern of abuse

Limitations of Regulation: How harmful content is pushed on platforms and regulation is not meeting the standard of prevention

Underestimating Impact and Extent of Harm: Limitations of record keeping which is related to the harms being underestimated

Introduction

The Centre for Protecting Women Online welcomes the opportunity to respond to Ofcom's consultation on the introduction of encouraging or assisting serious self-harm and cyberflashing as priority offences under the Online Safety Act. This response adopts a technology facilitated violence against women and girls lens. This means we

understand technology facilitated harms as part of a wider pattern of gendered violence that is shaped and amplified by digital platforms and rapidly emerging technologies.

We support the decision to make both offences priority harms. However, we are concerned that the proposed changes do not fully reflect how these harms affect women and girls, or how they are changing as a result of artificial intelligence. As a result, the proposals risk focusing too much on removing harmful content after it appears, rather than preventing harm from happening in the first place.

We also have concerns regarding the role and status of risk profiles and the register of risks within the proposed framework. As currently framed, it appears that platforms are required to take these into account when conducting their own risk assessments, rather than adhere to them as binding standards. This introduces a level of discretion that may weaken the consistency and effectiveness of risk mitigation. Such discretion is particularly concerning in light of recent and historic high profile examples of harm as a result of misuse of platforms. Greater clarity and stronger expectations may therefore be needed to ensure that risk is assessed and addressed in a consistent and evidence-based way across services.

We are also concerned that Ofcom's view that these changes will have minimal impact does not reflect the scale of the problem. Addressing technology facilitated violence against women and girls requires meaningful changes to how platforms are designed and operate. Small updates to existing processes will not be enough to reduce harm in practice.

In addition, the proposals should more explicitly account for the differing functions, capacities, and risk profiles of platforms of different sizes. While larger platforms may already have established safety mechanisms, it cannot be assumed that smaller or less visible services operate to the same standards. This highlights a broader challenge for enforcement and compliance, particularly in relation to platforms operating in less regulated or harder-to-reach areas of the internet. Evidence suggests that such services can fall outside the effective scope of regulatory frameworks that are primarily designed with larger, mainstream platforms in mind. This is particularly concerning given that some smaller and less regulated sites may be more likely to host or facilitate harmful content.

There is also a clear gap between the intention set out by Government and the approach taken in this consultation. Government statements linked to the cyberflashing offence emphasised that platforms should prevent harmful content from reaching users, particularly women and girls. In contrast, the proposals mainly extend existing measures and do not introduce strong new requirements to stop harm before it

occurs. This raises concerns about whether the regulatory framework will deliver the level of protection that was intended.

Suicide and self-harm

We understand why Ofcom proposes to combine suicide and self-harm into a single category. However, this approach risks overlooking important differences in how these harms affect different groups. Evidence from organisations such as YoungMinds and Refuge shows that self-harm content affecting girls and young women is often shared and reinforced within online peer groups. In these spaces, harmful behaviours may be normalised or presented as a way of coping, rather than recognised as a serious risk (YoungMinds, 2023; Refuge, 2022). This is different from suicide related content, which is more often linked to immediate crisis situations.

These harms are also shaped by inequality. Girls and young women, especially those who identify as LGBTQ+, disabled or from minoritised backgrounds, are more likely to experience both self-harm content and other forms of online abuse. For this reason, it is important that services are required to assess risk in a way that reflects gender and other compounding experiences. A single combined category should not lead to a one size fits all approach.

We welcome the updates to the Risk Profiles and Register of Risks, but the role of platform design needs to be clearer. In particular, recommender systems can push users towards more harmful content over time. Research by the Center for Countering Digital Hate shows how quickly users can be exposed to increasingly extreme self-harm material (2022). This is especially important for young women, who are more likely to engage with visual content and online communities where harmful behaviours can be amplified.

The consultation also does not go far enough in addressing the risks linked to AI. The current focus on user to user and search services may exclude some AI systems, especially those that involve one to one interaction. This creates a gap in the regulatory framework.

This is a serious concern because AI systems can generate harmful content directly. They can produce personalised and convincing responses that encourage self-harm (De Freitas et al., 2024). Research shows that these systems can reinforce harmful thinking over time (Weilnhammer et al., 2026), and that safety measures can be bypassed through techniques such as jailbreaking (Schoene and Canca, 2025). There is also evidence that AI systems do not reliably match the judgement of trained clinicians in assessing risk (McBain et al., 2025). A case reported in 2023 described a man in Belgium who died after extended interaction with an AI chatbot that encouraged his

suicidal thoughts (AI Incident Database, 2023). This shows that AI systems can actively contribute to harm, rather than simply making harmful content easier to find.

Cyberflashing

We support the inclusion of cyberflashing as a priority offence. However, it should be clearly recognised as a form of sexual violence against women and girls. Evidence from organisations such as Glitch and End Violence Against Women Coalition show that cyberflashing is commonly used to intimidate and harass women, and that it limits their ability to participate safely online (Glitch, 2021; End Violence Against Women Coalition, 2023).

The proposed measures do not yet meet the standard of prevention set out by Government. Tools such as blocking and muting only work after harm has already happened. They do not stop the initial exposure. Effective regulation should require platforms to prevent this content from being sent in the first place.

The risks of this offence are also increasing as a result of AI. Tools that generate sexual images can create non-consensual content quickly and at scale. This makes it easier to target women and girls and increases the overall level of harm.

We support the proposed updates to record keeping, but services should be required to collect better data. This should include information broken down by gender, as well as data on repeat harm and AI generated content. Without this, it will be difficult to understand whether the measures are working.

We are concerned that the overall impact of these harms is underestimated. Evidence from civil society shows that online abuse affects mental health and limits women's participation in public and digital life. This should be more clearly reflected in the impact assessment.

Conclusion

In conclusion, we support the introduction of these offences as priority harms. However, the regulatory approach needs to go much further. It should focus on preventing harm, not just responding to it. It should also recognise the role of AI in creating new risks. Most importantly, it should take a clear and consistent approach to technology facilitated violence against women and girls. Without these changes, the framework will not be strong enough to address the scale and nature of harm and will be a missed opportunity to truly protect women and girls.

Bibliography

AI Incident Database (2023) *AI-assisted suicide case report (Report 2864)*. Available at: https://incidentdatabase.ai/reports/2864/ (Accessed: 14 April 2026)

Center for Countering Digital Hate (2022) *Deadly by Design: Algorithmic amplification of self-harm and suicide content*. Washington, DC: CCDH.

De Freitas, J., Uğuralp, A.K., Oğuz-Uğuralp, Z. and Puntoni, S. (2024) 'Chatbots and mental health: Insights into the safety of generative AI', *Journal of Consumer Psychology*, 34(3), pp. 481–491.

End Violence Against Women Coalition (2023) *Technology-facilitated violence against women and girls*.

Glitch (2021) *The Ripple Effect: Online abuse and its impact on women and marginalised people*.

McBain, R.K., Cantor, J.H., Zhang, L.A., Baker, O., Zhang, F., Burnett, A. and Yu, H. (2025) 'Evaluation of alignment between large language models and expert clinicians in suicide risk assessment', *Psychiatric Services*, 76(11), pp. 944–950.

Pavón Pérez, Á., Farrell, T., De Kock, C., Jurasz, O., Nozza, D. and Fernandez, M. (2026) 'Still unsafe: What is holding us back on online safety for women', *AI and Ethics*. Advance online publication.

Refuge (2022) *The Online Safety Gap: The online abuse of women and girls*.

Schoene, A.M. and Canca, C. (2025) 'For argument's sake, show me how to harm myself: Jailbreaking LLMs in suicide and self-harm contexts', in *Proceedings of the IEEE International Symposium on Technology and Society (ISTAS)*. IEEE, pp. 1–7.

Weilnhammer, V., Hou, K.Y., Luettgau, L., Summerfield, C., Dolan, R. and Nour, M.M. (2026) 'Vulnerability-amplifying interaction loops: A systematic failure mode in AI chatbot mental-health interactions', *arXiv preprint*, arXiv:2602.01347.

YoungMinds (2023) *Self-harm and young people: Experiences and support needs*.