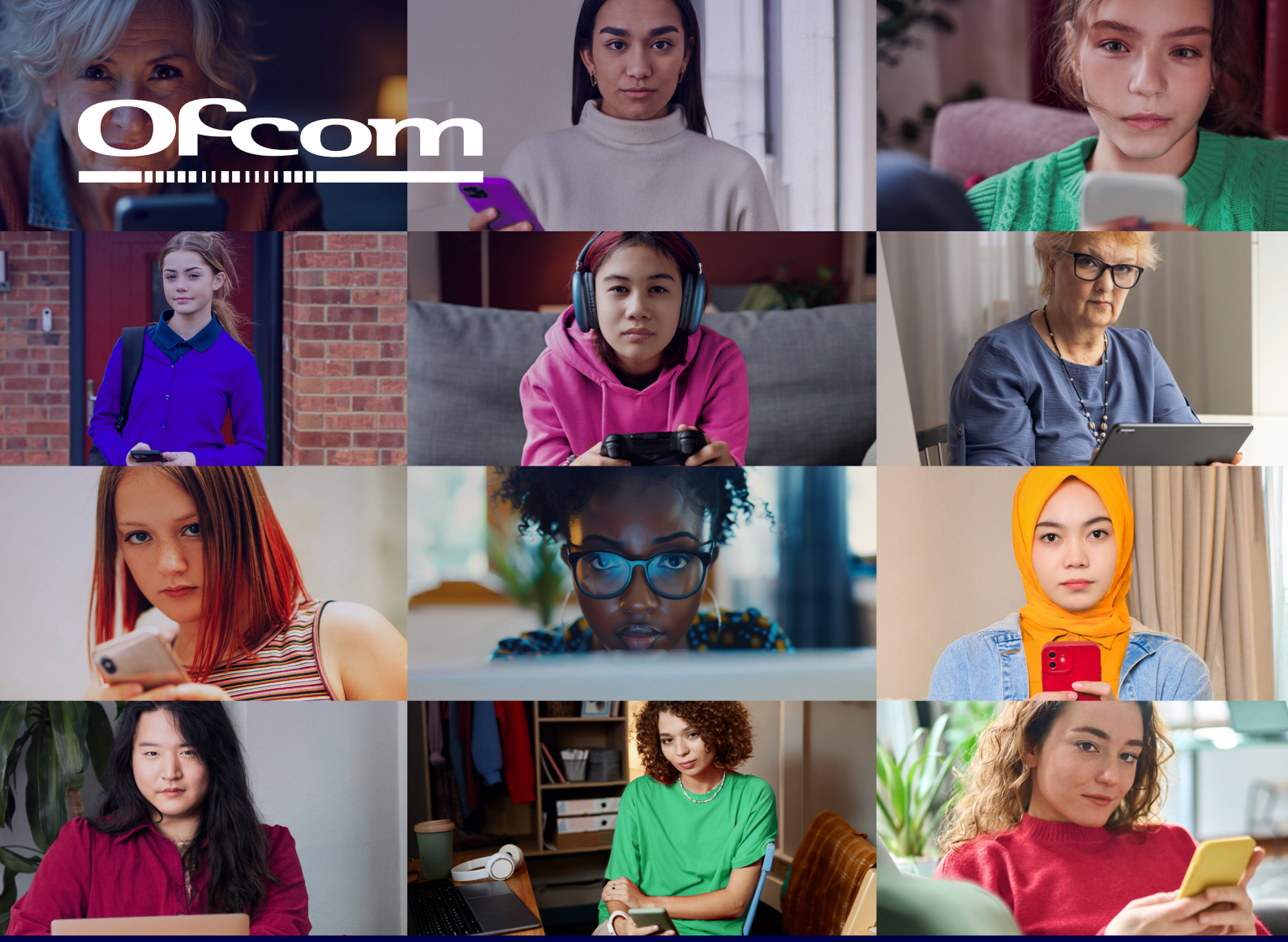

A Safer Life Online for Women and Girls

Practical Guidance for Tech Companies

Published 25 November 2025

Welsh version available

Contents

WARNING: THIS GUIDANCE CONTAINS CONTENT THAT MAY BE UPSETTING OR DISTRESSING.

Section

Annex

Foreword

When we speak to women and girls who have experienced online abuse, their stories stay with us. Survivors have told us how an intimate image shared without consent shattered their sense of safety. Journalists, politicians and athletes describe the relentless online abuse they face while simply doing their jobs.

As Ofcom wrote this Guidance, we listened closely to the frontline support workers, researchers, and advocates who have spent years supporting women and engaging with men and boys to prevent online gender-based harms. Their insights are clear: the digital world is failing to tackle harms to women and girls. It's failing to protect them from image-based sexual abuse, stalking and coercive control, pile-ons and coordinated harassment, and the insidious culture of misogyny that thrives online.

This Guidance sets out practical, ambitious steps for online services to make their platforms safer. Simple features such as nudges, warnings and limits on posts can help prevent perpetrators abusing women online. Ensuring that survivors and victims' voices are heard can help to build easier, simpler and better safety tools and reporting systems. Crucially, tech firms must take responsibility for safety from the outset, testing their products for how they can be misused before they roll them out, being more transparent about their users' experiences, and making sure their teams are properly resourced and trained.

At Ofcom we are very clear that safety must never come at the cost of silencing legitimate online debate. Freedom of expression is a cornerstone of democratic life, and it must be protected online. The right to speak, to challenge, to create, and to connect, without fear of abuse, is essential. Our goal is not to silence voices, but to ensure that digital spaces allow everyone to participate fully.

We will hold tech firms to account. Companies can expect to face enforcement action where they do not meet their legal duties. And we will shine the light of transparency, so it is clear to the public whether tech firms are making meaningful progress towards ending online harms that disproportionately affect women and girls.

1. Introduction

1.1 The Online Safety Act 2023 ('the Act') places clear requirements on online service providers ('service providers')[1] to address illegal content, which includes gender-based harms such as intimate image abuse, stalking and coercive control. The Act also requires service providers to protect children from other harmful content, such as gender-based harms like misogynistic abuse and sexual violence, and to shield them from pornographic content. We have published Codes of Practice and risk assessment guidance setting out how we expect service providers to tackle illegal content and protect children.[2]

1.2 Now that these duties are in force, Ofcom's role is to hold service providers to account, using our robust enforcement powers as needed. The foundational protections that are set out in our Codes and risk assessment guidance will benefit women and girls as well as all UK users of online services.

1.3 In addition, when passing the Act, Parliament included a requirement on Ofcom to produce additional, dedicated guidance on women and girls' online safety. This Guidance sets out how service providers, including dating apps, social media, gaming, pornography sites[3] and search services, can address content and activity that disproportionately affects women and girls.[4]

1.4 There is clear evidence that women and girls experience unique and serious risks online. They are more at risk of having their intimate images – including deepfakes – shared without their consent, they are disproportionately targeted by misogynistic abuse and pile-ons, and they are the main survivors and victims of stalking and coercive control.[5]

1.5 Throughout the development of this Guidance, we have heard from organisations on the front line tackling these harms, including services supporting survivors and victims, organisations working with men and boys, and experts researching how these harms manifest. Most importantly, we've spoken to survivors and victims who have told us that they want to see providers make meaningful, practical changes to their services, so they are safer for the millions of women and girls across the UK who use them daily.

---

[1] Such services are defined under Part 3 of the Act and include user-to-user and search services. Throughout this document, we refer to the online platforms themselves as 'services', and the legal entity that provides the service as a 'service provider' or 'provider'.

[2] We published our statement on Illegal Content Codes and Risk Assessment Guidance in December 2024, setting out how services must approach their new duties relating to illegal harms, and we published the Protection of Children Codes and Risk Assessment Guidance in April 2025.

[3] The Guidance does not apply to Part 5 services (providers that publish or display pornographic content themselves, with no user-to-user interactions or search content). Ofcom has produced separate guidance for these services.

[4] Section 54 of the Act says the Guidance must focus on 'content and activity….in relation to which such providers have duties set out in [Part 3] or Part 4 of the Act' and 'which disproportionately affects women and girls'.

[5] In Chapter 2 of the Guidance, we provide an overview of key evidence sources on harms to women and girls online. For a more detailed overview of how harms manifest online, including risks to women and girls, see our Illegal Harms Register of Risks, and our Children's Register of Risks.

1.6 This Guidance sets out nine actions for tackling online gender-based harms. We illustrate these actions with practical examples of the measures set out in our Codes and guidance on Illegal Harms[6], Protection of Children[7] and Transparency[8] to mitigate gender-based harms. These 'foundational steps' bring together relevant measures outlined in our Codes of Practice which outline how providers can comply with the enforceable duties under the Act related to illegal content and protection of children. We are ready to take enforcement action if providers do not comply with their duties.

1.7 This Guidance also goes further. We also illustrate the actions with additional good practice – features, tools and processes – that providers can take to deliver ambitious and meaningful changes towards a safer life online for women and girls and all those impacted by online gender-based harms. While some of these harms are not illegal and providers must also ensure their users are still able to express themselves freely, the Guidance reflects the real experiences of women and girls and gives voice to their calls for better protections and more choice from the online services they use.

1.8 We call on providers to prioritise women and girls' safety. While the Guidance focuses on harms disproportionately affecting women and girls, doing so will help raise safety outcomes for anyone who may be affected by the harms. This will create a safer life online for everyone and ensure that all users can benefit from the opportunities of actively participating in online services. We plan to closely monitor and shine a light on how service providers choose to protect their users, including how they have applied this Guidance, to help users make informed decisions about their online experiences.

1.9 In addition to Ofcom's other duties,[9] Ofcom must carry out its functions compatibly with fundamental rights in accordance with its duties under the Human Rights Act 1998, including the right to freedom of expression under Article 10 of the European Convention on Human Rights.[10] We have carefully considered the rights protected by the Convention when preparing this Guidance.[11] We are satisfied that the Guidance is proportionate to the legitimate aim of protecting women and girls, and people in the UK more generally, from harm online.

---

[6] See the Illegal Content Codes and the Illegal Content Judgements Guidance.
[7] See the Protection of Children Codes and the Guidance on Content Harmful to Children.
[8] See the Transparency Reporting Statement.
[9] These include Ofcom's general duties under section 3 of the Communications Act 2003 (CA 2003), duty to carry out and publish an impact assessment under section 7 of the CA 2003, Media literacy duties in section 11 of the CA 2003, duties under section 149 of the Equality Act 2010 and section 75 of the Northern Ireland Act 1998, and duties in relation to Welsh language standards. In preparing the Guidance, we have also had regard to the Government's Statement of Strategic Priorities for Online Safety.
[10] Article 10(1) states that the right to freedom of expression includes the freedom to hold opinions and to receive and impart information and ideas without interference by public authority. Article 10(2) states that this right may be restricted in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary. Any limitation on the right to freedom of expression must be prescribed by law, pursue a legitimate aim and be necessary in a democratic society. In order to be 'necessary', the restriction must correspond to a pressing social need, and it must be proportionate to the legitimate aim pursued.
[11] In relation to the foundational steps in the Guidance, we also gave careful consideration to human rights impacts of our Codes and guidance when preparing our Illegal Harms and Protection of Children statements.

1.10 This Guidance covers the following:

- **Chapter 1** introduces concepts used throughout the Guidance.
- **Chapter 2** explores how harms disproportionately affecting women and girls manifest online, and the severe and wide-ranging impacts they have, including on women and girls' ability to freely express themselves and safely participate in life online.
- **Chapter 3**, **Chapter 4**, and **Chapter 5** set out nine action areas for online service providers to tackle these harms.
- **Annex 1** includes a glossary with definitions of the terms we have used throughout the Guidance.

What harms do we focus on in this Guidance?

1.1 Our Guidance focusses on content and activity where service providers have duties under the Act and which disproportionately affects women and girls. Addressing this type of content and activity in the round requires an understanding of how gendered harms overlap and how they cut across illegal content and activity,[12] and content and activity harmful to children.[13]

1.2 This content and activity include the harmful ways online services are used to control, exploit, monitor, silence, humiliate, abuse, and threaten women and girls because of their gender. Research shows that those with multiple protected characteristics, such as LGBTQ+ communities and women from ethnic minority backgrounds, face additional risks of harm.[14]

1.3 The impacts are severe and wide-ranging, from inhibiting the safety and participation of women and girls in online spaces – ultimately affecting their ability to engage and express themselves freely online – to the normalisation of misogynistic attitudes and behaviours. Ofcom's research found that women are significantly less likely than men to believe that the benefits of being online outweigh the risks, and to say they can share their opinions and have a voice online. Women and girls are also more likely to be adversely impacted by potential harmful content online. Evidence also shows that the way providers operate and monetise content can push men and boys towards content that normalises, glorifies, and

[12] For a full list of priority offences covered under the Act, and which may count as illegal content, see Overview of Illegal Harms which defines and categorises over 130 'priority offences'. See our Illegal Harms Register of Risks for detailed discussions of on how illegal harms manifest and the wider impacts of these harms, including how certain kinds of harm such as intimate image abuse and controlling and coercive behaviour disproportionately affect women and girls.

[13] For a full list of content set out under the Act which is harmful to children, see Guidance on Content Harmful to Children. There are three types of content specified by the Act, some of which disproportionately affect women and girls. Primary priority content refers to pornographic content, suicide content, self-harm content and eating disorder content. Priority content refers to a range of harms including abuse content, content inciting hate, content encouraging violence, and bullying content. Non-designated content refers to content, not included in the Act, that presents a material risk of harm to an appreciable number of children. Under the Act, services are required to prevent children from encountering primary priority content that is harmful to children, and to protect children in age groups judged to be at risk of harm from priority content and non-designated content. For detailed discussions of evidence on content harmful to children, see our Children's Register of Risks.

[14] The Global Partnership, 2023. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 3 November 2025].

encourages misogynistic abuse and sexual violence.[15] Online cultures that promote actions and attitudes that demean, oppress and otherwise harm women and girls have a broader influence on attitudes and behaviours both online and in the offline world.

1.4 For this Guidance, we focus on four overlapping forms of harm from content and activity covered by the duties under the Act and which disproportionately affect women and girls. We refer to these collectively as **online gender-based harms**:[16]

- **Misogynistic abuse and sexual violence:** This describes the circulation – or promotion – of abusive, hateful and violent content that actively encourages or cements misogynistic ideas or behaviours, including through the normalisation of sexual violence. This includes both illegal content (e.g. illegal threats and abuse) and content harmful to children (e.g. abusive, hateful and violent content), which is often algorithmically amplified and targeted towards young men and boys.
- **Pile-ons and coordinated harassment**: This describes cases where groups of coordinated perpetrators target a specific woman or girl, or groups of women and girls with abuse, hate and threats of violence. As with misogynistic abuse and sexual violence, this can include both illegal content (e.g. threats and intimate image abuse) and content harmful to children (e.g. abusive, hateful and violent content). While pile-ons can happen to any user, they often target women in public life, such as journalists and politicians.
- **Stalking and coercive control**: This describes criminal offences that involve using technology for (a) stalking; or (b) coercive and controlling behaviour in the context of an intimate or family relationship.
- **Image-based sexual abuse**: This refers to the criminal offences of (a) intimate image abuse (the non-consensual sharing of intimate images) and (b) cyberflashing (sending explicit images to someone without their consent). We also cover self-generated indecent imagery.[17]

1.5 While these harms disproportionately affect women and girls, they do not do so exclusively. Anyone can be targeted by pile-ons, intimate image abuse, stalking and coercive control. Furthermore, abusive misogynistic content is often pushed towards boys – stakeholder feedback highlighted that boys often find this content upsetting as it can normalise harmful narratives and beliefs of masculinity that affects them.[18] We provide further detail on how these harms manifest and their impacts in Chapter 2.

---

[15] Women's Aid, 2023. Influencers and Attitudes: How will the next generation understand domestic abuse?. [accessed 29 October 2025]; Vera-Gray, F., McGlynn, C., Kureshi, I., Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology*, 61 (5). [accessed 29 October 2025]; Vodafone, 2024. AI 'Aggro-rithms': young boys are served harmful content within 60 seconds of being online. [accessed 29 October 2025].

[16] This is the term we use to describe our approach to 'content and activity that disproportionately affects women and girls' as set out in section 54 of the Act. There are other commonly used terms to describe these harms, including online violence against women and girls (VAWG) and tech-facilitated gender-based violence. While these other terms refer to a similar subset of harms, we have selected online gender-based harms as it best reflects the scope of the Guidance.

[17] Self-generated indecent imagery (SGII) refers to sexual imagery created by a child of themselves. All SGII is classified as child sexual abuse material and is therefore illegal.

[18] Ofcom / Men and Boys Roundtable, 29 May 2025.

1.6 As noted above, we have carefully considered the implications of this Guidance, specifically the identification of a broad range of content, on users' rights and fundamental freedoms, including freedom of expression. We have taken steps to ensure this Guidance is proportionate to the risk of harm, taking into consideration the risk of harm that can manifest offline. In line with this consideration, it is important to note that we do not expect service providers to remove or heavily monitor all the content and activity captured under the broad category of harm. However, we consider it right for providers to adopt a holistic approach when examining the impact of their services on the safety of women and girls, including when making decisions about the policies, tools, and features they offer users, and how those choices may affect women and girls' safety.

What actions are we asking service providers to take?

1.11 The Act emphasises a safety-by-design approach for providers to reduce risks on their services from the outset.[19] This includes making improvements to existing systems or features on longstanding services, as well as ensuring new services or features are designed with safety in mind from the outset.

1.7 Using this safety-by-design approach, we set out nine actions providers can take to address the challenge of online gender-based-harms across each stage of operating and designing their service.[20] Figure 1 sets out the nine actions, which are split across Chapter 3, Chapter 4, and Chapter 5:

[19] See, for example, section 1(3) of the Act which explains that the duties imposed on providers in the Act seek to secure (among other things) that services regulated by this Act are safe by design.

[20] Our approach draws on previous work which uses similar concepts to demonstrate how to address online gender-based harms and online safety more generally, including academic literature on safety-by-design and a report by the eSafety Commissioner on tackling 'gendered violence' through safety-by-design. [accessed 3 November 2025].

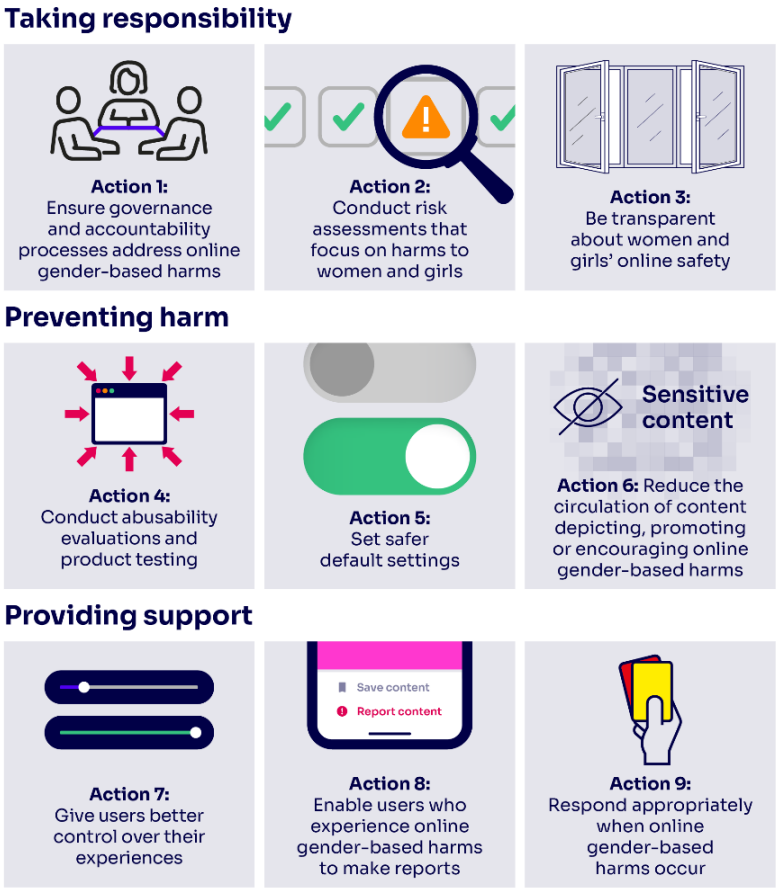

Figure 1: Nine actions areas in Chapter 3, Chapter 4, and Chapter 5

1.12 For each action, we first explain where we have already set out the measures providers can take to meet their duties under the Act related to illegal content and protection of children. This information is taken from the Codes and risk assessment guidance we have published across our work on illegal harms, protection of children, and (only as applicable to a smaller number of providers) transparency. We refer to these collectively as **'foundational steps**.' These steps are wide-ranging and cover areas including conducting risk assessments,[21] transparency reporting[22] and Codes related to improving user safety.[23] For some of these foundational steps, we include **case studies** to showcase what taking action could look like in a particular context, and how the foundational steps will address harms to women and girls. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document.

[21] Our risk assessment guidance is intended to assist services in fulfilling their legal obligations under the Act related to risk assessments. It does not represent a set of compulsory steps that services must take. The Illegal Content Risk Assessment Guidance was published in December 2024. The Children's Risk Assessment Guidance was published in April 2025.

[22] This applies to categorised services only. Section 78(1) Act requires Ofcom to produce guidance for such services about the transparency reporting framework, and we have published guidance in accordance with that duty.

[23] Codes describe measures recommended for the purpose of compliance with duties. If service providers implement measures recommended in Codes, services will be treated as complying with the relevant duties. However, the Act allows service providers to adopt alternative measures, provided they keep a record of what they have done and explain how they think the relevant safety duties have been met. The Illegal Content Codes came into force in March 2025. The Protection of Children Codes came into force in July 2025.

1.13 Second, we include '**good practice steps'** which set out practical ways in which providers can go further to demonstrate a commitment to taking responsibility for user safety and preventing and responding to online gender-based harms.

1.14 These good practice steps draw on a range of evidence sources including workshops we ran with a wide range of stakeholders in late 2024, as well as evidence gathered through the consultation process. We have also drawn on academic research, expert reports, industry practice and our work under our media literacy duties.[24] For some of these good practice steps, we also include case studies to showcase what taking action could look like in a particular context and showcase how providers can address specific harm areas like misogynistic abuse or intimate image abuse.

1.15 While the foundational steps are connected to the duties providers must comply with under the Act in relation to illegal content and protection of children, the good practice steps show practical and feasible ways to go further. We call on providers to prioritise a holistic response to misogynistic abuse and sexual violence, pile-ons and coordinated harassment, stalking and coercive control, and image-based sexual abuse.

1.16 When taking any steps, providers should also note the need to comply with any relevant data protection laws. Providers remain accountable for their obligations under data protection law when implementing any of the steps set out in the Guidance.[25]

What does this mean in practice?

1.17 The Guidance sets out how service providers can tackle content and activity that disproportionately affects women and girls. To do this, we include a number of steps designed to address online gender-based harms. These range from undertaking better risk assessments, to introducing stronger defaults, more supportive reporting and better user controls for everyone on the service. Where providers take these steps, we expect the standard of safety will rise for anyone using the service, and anyone targeted or impacted by these harms.

1.18 The good practice steps set out in this Guidance are non-exhaustive. Technology evolves rapidly and there are likely to be ambitious, effective and innovative interventions to secure the safety of women and girls which we have not included or could emerge as safety technology develops. We would encourage service providers to explore these new and emerging solutions and, where possible, introduce them into their systems and processes.

1.19 Harms evolve rapidly, and we expect providers to regularly look at what they may need to do to respond to changing threats and risks from online gender-based harms. We also expect services with the highest risk and largest reach to do more to ensure they achieve safer experiences for women and girls.

---

[24] Ofcom has a statutory duty to promote media literacy and to carry out research into media literacy under section 11 of the Communications Act 2003, including specific work related to content and activity that disproportionately affects women and girls. Where relevant, we draw on this duty and our media literacy work to illustrate good practice for online services. We define media literacy as "the ability to use, understand and create media and communications across multiple formats and services".

[25] The Information Commissioner's Office's ('ICO') guide to UK GDPR offers guidance on carefully assessing the necessity and proportionality of personal information processing and determining the minimum amount of personal information required to achieve the intended purpose. Providers will also need to ensure that they have a lawful basis for processing personal data under Article 6 UK GDPR. See the ICO's guidance on lawful basis.

1.20 Finally, we do not expect all service providers will need to – or will be able to – introduce all of the good practice steps we have set out under each action. Some steps may only be relevant or applicable to certain services (for example because of the user base size, risk level, user demographics, the design of the service or business model) and we note this throughout. For example, the way online gender-based harms manifest on search services would be different to how they may manifest on user-to-user services, including social media, dating, gaming, porn, file-sharing and messaging services.

2. What are online gender-based harms?

2.1 This chapter explains our understanding of online gender-based harms. We set out how these harms relate to illegal content and content harmful to children under the Act.[26] We outline evidence about how online gender-based harms manifest and the different tactics used by perpetrators of these harms. We set out how the good practice steps in this Guidance can help providers tackle these harms on their services.

2.2 For further information on the causes and impacts of the specific harms referenced, see the Illegal Harms Register of Risks and Children's Register of Risks. For further information on how we understand what is in and out of scope of these harms, see the Illegal Content Judgements Guidance and the Guidance on Content Harmful to Children.

Background

2.3 **Online gender-based harms are systemic and intersectional**. They recreate and amplify existing gender discrimination, sometimes introducing new forms of harm.[27] Evidence indicates that women and girls with additional factors including age, race, ethnicity, socio-economic status, sexual orientation, gender identity, disability and religion experience heightened and specific risks and impacts of online gender-based harms.[28] [29]

2.4 **Online and offline harms often co-occur and overlap.** This is especially true for forms of online gender-based harms that include patterns of behaviour. The National Stalking Helpline reported that all cases presenting to the helpline include online elements.[30] In addition, different forms of online gender-based harms overlap with one another. For example, intimate image abuse can be a form of coercive control.[31]

2.5 **Online gender-based harms are exacerbated by societal norms such as victim-blaming**. Women and girls who experience abuse are often told they should have done more to keep themselves safe and can be blamed for being online. This results in an unnecessary burden

[26] For a full list of illegal content set out under the Act, see Overview of Illegal Harms which defines and categorises over 130 'priority offences'. For a full list of content harmful to children set out under the Act, see Guidance on Content Harmful to Children.
[27] The Global Partnership, 2023. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 29 October 2025].
[28] Crenshaw, K., 2013. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, *University of Chicago Legal Forum*, 1989 (1). [accessed 29 October 2025]; Gray, K. L., 2011. INTERSECTING OPPRESSIONS AND ONLINE COMMUNITIES: Examining the experiences of women of color in Xbox Live. *Information, Communication & Society,* 15 (3). [accessed 18 August 2025].
[29] Department for Science, Innovation & Technology and PUBLIC, 2025. Platform Design and the Risk of Online Violence Against Women and Girls. [accessed 29 October 2025]; The Global Partnership, 2023. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 29 October 2025]; UNESCO, 2023. Exposing Technology-Facilitated Gender-Based Violence in an Era of Generative AI. [accessed 24 September 2025].
[30] Suzy Lamplugh Trust, 2021. Unmasking Stalking: A Changing Landscape. [accessed 11 August 2025].
[31] Refuge, 2020. The Naked Threat. [accessed 30 September 2025].

on women and girls to avoid, respond to and cope with gender-based harms.[32] This is known as 'safety work'[33] and hinders their ability to participate freely in online life.

2.6 **Online gender-based harms can curtail women and girls' human rights**. Experiencing online gender-based harms may impact the right of women and girls to live free from violence and fear and can also impact their right to privacy. Abuse, or the fear of abuse, can limit the ability of women and girls to express themselves freely and participate in public and political life.

Online gender-based harms

2.7 The recommendations in this Guidance focus on addressing four categories of online gender-based harms, which evidence suggests disproportionately affect women and girls:

a) Misogynistic abuse and sexual violence;
b) Pile-ons and coordinated harassment;
c) Stalking and coercive control; and
d) Image-based sexual abuse, including intimate image abuse, self-generated indecent imagery and cyberflashing.

2.8 These four categories do not cover all types of harmful content and activity that may impact women and girls' experiences online,[34] such as grooming for the purposes of child sexual exploitation and abuse (CSEA), modern slavery and human trafficking, and eating disorder content. Those harms are either priority offences for illegal harms or primary priority content for protection of children under the Act. We set out our recommendations for how service providers can assess and mitigate the risk of these harms in our Codes and guidance related to those duties.

2.9 Online gender-based harms are constantly evolving alongside technological developments. This means that new harms, and new manifestations of existing harms, are always emerging. This Guidance sets out high-level actions providers can take to tackle online gender-based harms regardless of perpetrator tactics or technologies.

2.10 It can be challenging for providers to identify online gender-based harms. Providers should consider additional context available about relevant content and activity, and how it is presented on the service. This could include the use of code words, whether there is a wider pattern of behaviour present, and, where appropriate, information about the perpetrator.

---

[32] Vera-Gray, F. and Kelly, L., 2020. Contested gendered space: public sexual harassment and women's safety work, *International Journal of Comparative and Applied Criminal Justice*, 44 (4). [accessed 29 October 2025].
[33] Gillett, R., 2021. "This is not a nice safe space": investigating women's safety work on Tinder, *Feminist Media Studies*, 23 (1). [accessed 29 October 2025].
[34] We recognise that some forms of child sexual exploitation and abuse (CSEA) disproportionately affect girls. For instance, girls are at a greater risk of experiencing grooming and being depicted in child sexual abuse material (CSAM). We are not focusing on CSEA in this guidance, but we do highlight CSEA measures set out in the Illegal Content Code for user-to-user services and search services as they apply to girls. We also recognise that some of the ways that CSEA manifests online, such as self-generated indecent imagery, overlap with online gender-based violence. For more information see Section 2 of the Illegal Harms Register of Risks.

Our approach

2.11 This Guidance focuses on the four categories of online gender-based harms listed above. For each of these harm areas, we engaged with both qualitative and quantitative evidence to understand the disproportionate and distinct effect the harm has on women and girls. This could be on the basis that women and girls are more likely to experience the harm, that the harm has a disproportionate effect on women and girls, or both.

2.12 However, there are some limitations with the research and data available on online gender-based harms. First, under-reporting is likely across all online gender-based harms, because certain groups, including men and boys and women and girls with additional characteristics, face specific barriers and social stigma around reporting. This can impact our understanding of how widespread online gender-based harms are and who experiences them. Second, the landscape and terminology of online gender-based harms has developed over time. This means that the evidence base addressing how these harms manifest does not always have consistent definitions, making robust cross-comparison challenging.[35]

2.13 We also consider which harms have a distinct effect on women and girls. For the purposes of this Guidance, we look at evidence about the effect of online gender-based harms on the normalisation of misogynistic abuse and sexual violence, and the distinct experiences of women in public life targeted by pile-ons.

2.14 None of the online gender-based harms we focus on exclusively affect women and girls. Anyone may encounter these harms or be a target for perpetrators. The good practice steps set out in this Guidance will therefore reduce harm more widely, by improving how service providers take responsibility for user safety, prevent these harms from occurring, and respond to them when they do.

2.15 In examining disproportionate effects on women and girls, we consider how multiple characteristics can combine to increase people's vulnerability to online gender-based harms. For example, women politicians from marginalised groups are more likely to be targeted by perpetrators of pile-ons because of their race or ethnicity[36] and services' detection and response to this can also be affected by overlapping forms of discrimination.[37] Online gender-based harms can also have a disproportionate effect on men and boys with additional protected characteristics. Transgender[38] and non-binary people[39] experience heightened risks of online gender-based harms. Depending on the sample size of the research, the impact of intersecting characteristics cannot always be

---

[35] King's Global Institute for Women's Leadership (Schmid, C., Fearnside, H. and Rohregger, N.), 2024. Measuring Gender Equality in the UK: Data on Violence Against Women and Girls. [accessed 10 October 2025].

[36] In a study by Amnesty International, Black, Asian and Minority Ethnic (BAME) women MPs received almost half (41%) of the abusive tweets in the run-up to the 2017 election, despite there being almost eight times as many white MPs in the study. Source: Amnesty International UK, 2017. Black and Asian Women MPs Abused More Online. [accessed 29 October 2025].

[37] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S. and Kaufmann, Z.), 2021. Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online. [accessed 29 October 2025].

[38] Responses to our February 2025 consultation: Galop, p.1-4; The Minderoo Centre for Technology and Democracy at the University of Cambridge, p.2; NSPCC, p.7-8.

[39] Evidence suggests non-binary people have distinct online experiences, though small sample sizes limit cross-group comparisons. In Ofcom's wider research on UK internet users, 62 respondents aged 18+ identified as non-binary. Findings show they were more likely than men aged 18+ to encounter potentially harmful content: objectifying or demeaning portrayals of women (51% vs. 17%), group shaming based on views (35% vs. 13%), and cyberstalking or harassment (13% vs. 5%). Source: Ofcom, 2025. Online Experiences Tracker – Wave 8.

identified through quantitative research. However, throughout this Guidance we have tried to highlight evidence about where our key harm areas may disproportionately affect those with additional characteristics.

Misogynistic abuse and sexual violence

2.16 Within this section on misogynistic abuse and sexual violence we focus on abuse, violence and hatred targeted at women and girls as a group. This covers threats and abuse, addressed in the Illegal Content Judgements Guidance. It also covers the following kinds of content harmful to children, addressed in the Guidance on Content Harmful to Children: abuse and hate content and violent content. We also cover content depicting sexual violence, including certain types of pornographic content.

2.17 Misogynistic abuse and sexual violence describes content and activity which normalises, encourages or reinforces attitudes and behaviours which promote abuse, violence or hatred targeted at women and girls. Misogynistic abuse and sexual violence is not always targeted at a specific woman or girl but can include harmful and discriminatory ideas about women and girls in general or the glamourisation of violence against them, such as a post which expresses the view that women and girls should be subjugated by men or a comment which glorifies coercive and controlling behaviour.

2.18 By definition, misogynistic abuse has a distinct effect on women and girls because they are the group being targeted. They are more likely to see or experience content or language online which objectifies, demeans or otherwise negatively portrays women and they are more likely to be very concerned about this harm.[40] However, evidence shows that men and boys are also exposed to, and are negatively impacted by, this content.[41]

2.19 Evidence suggests that misogynistic beliefs that begin online can lead to offline violence, in both public and private spaces.[42]

Misogynistic abuse

2.20 Misogynistic abuse is perpetrated and witnessed in a variety of online spaces, across both larger services serving many audiences, and smaller services and communities dedicated to propagating misogynistic attitudes and behaviours.

2.21 On larger services, misogynistic abuse can include content that incites hatred towards women and normalises harmful ideas about how men and boys should act towards women and girls, including by commending abuse.

[40] UK women internet users aged 18+ are more likely than UK men internet users aged 18+ to say they have seen or experienced something online in the last four weeks which objectifies, demeans or otherwise negatively portrays women (23% of women users, compared to 17% of men users). Women internet users are also more likely than men internet users to say they are very concerned by this content (60% of women users, compared to 38% of men users). Source: Ofcom, 2025. Online Experiences Tracker – Wave 8.

[41] Ofcom, 2025. Online Experiences Tracker – Wave 8.

[42] Institute for Strategic Dialogue (Bundtzen, S.), 2023. Misogynistic Pathways to Radicalisation: Recommended Measures for Platforms to Assess and Mitigate Online Gender-Based Violence. [accessed 31 October 2025].

2.22 Much of the most popular content is produced by users with large followings. The content is framed as entertainment, aligning with interests such as self-improvement or gaming, and using formats such as memes and inspirational stories.[43]

2.23 Evidence highlights the influence of misogynistic content creators, or 'influencers', on the rise of misogyny in schools in the United Kingdom.[44] Boys are significantly more likely than girls to have viewed content from such influencers, and are significantly more likely to have a positive view of the content they produce.[45] This consistent exposure to misogynistic abuse normalises harmful attitudes towards women and girls and creates unhealthy perceptions of relationships, as well as a distorted view of what it means to be a successful and attractive man. For example, children and young people who engage with this content are almost five times more likely to agree with the statement that "hurting someone physically is okay if you say sorry after".[46]

2.24 Services and individual content creators may benefit financially from content that is more 'shocking' or 'extreme'. Evidence shows that recommender systems reward influencers who create content that promotes misogynistic abuse with greater reach, particularly to boys and young men.[47] This happens because algorithms are optimised for high engagement, which can incentivise the production of, and exposure to, polarising and harmful content.[48] These effects are exacerbated by service design features such as endless feeds and autoplay settings, since users are less likely to choose the next piece of content they engage with. For example, research has found that young people searching for friends, advice or shared groups are served content that is increasingly misogynistic through their recommender feeds.[49]

2.25 In this Guidance, we highlight good practice steps service providers can take to reduce the circulation and visibility of harmful content such as misogynistic abuse. These include deprioritising harmful content in recommender algorithms and de-monetising user-generated content which promotes online gender-based harms.

---

[43] Regehr, K., Shaughnessy, C., Zhao, M. and Shaughnessy, N., 2024. Safer Scrolling: algorithms popularise and gamify online hate and misogyny for young people. [accessed 20 August 2025]; Women's Aid, 2023. Influencers and Attitudes: How will the next generation understand domestic abuse?. [accessed 29 October 2025].

[44] Das NETTZ, Textgain and Federal Association for Countering Online Hate, 2024. Tracing Online Misogyny: An analysis of misogynist ideologies and practices from a German-international perspective. [accessed 3 November 2025].

[45] Internet Matters, 2023. "It's really easy to go down that path": Young people's experiences of online misogyny and image-based abuse. [accessed 29 October 2025].

[46] 19% of participants aged 7-18-years-old who had been exposed to content from misogynistic influencers thought "hurting someone physically is okay if you say sorry after hurting them", compared to 4% of those not exposed. Source: Women's Aid, 2023. Influencers and Attitudes: How will the next generation understand domestic abuse?. [accessed 29 October 2025].

[47] Vodafone, 2024. AI 'Aggro-rithms': young boys are served harmful content within 60 seconds of being online. [accessed 29 October 2025].

[48] Ofcom, 2025. Children's Register of Risks.

[49] Internet Matters, 2023. "It's really easy to go down that path": Young people's experiences of online misogyny and image-based abuse. [accessed 29 October 2025]; Vodafone, 2024. AI 'Aggro-rithms': young boys are served harmful content within 60 seconds of being online. [accessed 29 October 2025].

2.26 Misogynistic abuse can also be perpetrated within smaller, dedicated services or communities. Young people who are lonely, isolated or who have mental health concerns can be drawn into more radical and misogynistic communities.[50] Though they vary in size, functionality and organisation, such communities are alike in their promotion, imagining and organisation of highly misogynistic attitudes and behaviours, often overlapping with other discriminatory[51] and extremist views.

2.27 Perpetrators in misogynistic communities often use images as part of their abuse. This can include the creation and sharing of 'semen images'[52] or images depicting intimate scenarios based on specific cultural or religious contexts.[53] Abuse against LGBT+ people can also include images which disclose a person's sexuality leading to 'outing' and deepfake images which 'de-transition' transgender people.[54]

2.28 Some of these communities promote user anonymity through community norms or service functionalities. Some communities, such as 'incel' groups, use coded language and symbols which are not widely recognised or understood.[55]

2.29 The closed and dedicated nature of these communities means that some of the good practice steps in this Guidance, such as enabling user reports, may not be effective in these contexts. However, service providers can still address these communities, for example by conducting risk assessments and responding appropriately when harm occurs.

Sexual violence

2.30 A common type of violent content targeting women and girls is content which depicts, invokes, encourages or normalises sexual violence. Sexual violence disproportionately targets women and online depictions of sexual violence are a manifestation, and reinforcement, of existing patterns of offline violence.[56]

2.31 This can include certain types of extreme pornography which is illegal content.[57] The depiction of any act of strangulation in pornography is also to be made illegal.[58] However,

---

[50] Das NETTZ, Textgain and Federal Association for Countering Online Hate, 2024. Tracing Online Misogyny: An analysis of misogynist ideologies and practices from a German-international perspective. [accessed 3 November 2025]; Griffin, J., 2021. Incels: Inside a dark world of online hate, BBC News, 13 August. [accessed 29 October 2025]; Vodafone, 2024. AI 'Aggro-rithms': young boys are served harmful content within 60 seconds of being online. [accessed 29 October 2025].

[51] Das NETTZ, Textgain and Federal Association for Countering Online Hate, 2024. Tracing Online Misogyny: An analysis of misogynist ideologies and practices from a German-international perspective. [accessed 3 November 2025].

[52] Images or videos where semen is depicted on top of a non-intimate image.

[53] Law Commission, 2022. Intimate image abuse: a final report. [accessed 28 October 2025].

[54] Response to our February 2025 consultation: Galop, p.4.

[55] Moonshot, 2021. Incels: A Guide to Symbols and Terminology. [accessed 25 September 2025].

[56] The Crime Survey for England and Wales 2022 found that women are more likely to experience sexual assault (27% of women aged 16+ had experienced sexual assault, including attempts, since the age of 16 compared to 5.7% of men aged 16+). Source: Office for National Statistics (ONS), Crime Survey for England and Wales 2022. [accessed 16 October 2025].

[57] Extreme pornography describes a category of illegal sexual material which includes content that depicts non-consensual behaviours, physical violence and threats to life. For more information, see the Illegal Harms Register of Risks.

[58] The Government announced that the depiction of strangulation in pornography will be designated as a priority offence under the Act. Source: Ministry of Justice and Department for Science, Innovation and Technology, 2025. New laws to target online abuse and pornography. [accessed 13 October 2025].

content depicting sexual violence can include certain types of legal pornography.[59] Evidence shows that content depicting sexual violence can be found on both mainstream and less popular pornography sites.[60]

2.32 'Sexual scripts' refers to the sexual behaviours and attitudes an individual perceives or understands as pleasurable or acceptable, as a result of their social environment. Research suggests that consuming pornographic content contributes to the sexual scripts held by the viewer.[61] Research shows that content depicting sexual violence contributes to sexual scripts which normalise harmful and coercive behaviour and can lead to users being less likely to understand what sexual acts are unlawful or harmful. This in turn can normalise harmful behaviour, such as strangulation,[62] and contribute to sexual violence becoming more socially acceptable or going unreported.[63] It can also affect attitudes towards consent, as it is often implied (rather than discussed) in pornographic content.[64]

2.33 Furthermore, recommender systems may also increase the risk of users being exposed to content depicting sexual violence. Evidence shows that sexually violent content is recommended even to first-time viewers of pornography sites[65] and a research study found that 12% of analysable content on the landing pages of the top three user-to-user adult services in the UK described sexual activity that constituted sexual violence.[66]

2.34 In this Guidance, we explore good practice steps with particular relevance to services which allow pornographic content. For example, service providers which allow pornographic content should be aware that such services are at heightened risk of hosting sexually violent content. We set out how providers can consult with subject-matter experts to consider how to address sexually violent content.

---

[59] Pornographic content is Primary Priority Content which children must be protected from encountering.
[60] Bertin, G., 2025. Creating a Safer World – the Challenge of Regulating Online Pornography. [accessed 26 August 2025].
[61] Most evidence on the link between pornography and sexual scripts focuses on young men. The sample for the research study included 462 college-aged men based in the US. Source: Marshall, E., Miller, H. and Bouffard, J., 2018. Bridging the Theoretical Gap: Using Sexual Script Theory to Explain the Relationship Between Pornography Use and Sexual Coercion, *Journal of Interpersonal Violence,* 36 (9-10). [accessed 29 October 2025].
[62] Bertin, G., 2025. Creating a Safer World – the Challenge of Regulating Online Pornography. [accessed 26 August 2025].
[63] Researchers have argued that the availability of extreme pornographic content – including rape and non-consensual sexual penetration – sustains a culture in which sexual violence is not taken seriously and risks being normalised. Source: Vera-Gray, F., McGlynn, C., Kureshi, I. and Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology*, 61 (5). [accessed 29 October 2025].
[64] BBFC and Revealing Reality, 2020. Young people, Pornography & Age-verification. [accessed 29 October 2025].
[65] Vera-Gray, F., McGlynn, C., Kureshi, I. and Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology*, 61 (5). [accessed 29 October 2025].
[66] For their study, the authors used the World Health Organisation definition of sexual violence, which is likely to include both extreme pornographic content and legal pornographic content. They focused on four broad categories of sexual violence: sexual activity between family members; aggression and assault; image-based sexual abuse; and coercive and exploitative sexual activity. The analysis was carried out over a six-month period during 2017-2018. Source: Vera-Gray, F., McGlynn, C., Kureshi, I. and Butterby, K., 2021. Sexual violence as a sexual script in mainstream online pornography, *The British Journal of Criminology,* 61 (5). [accessed 29 October 2025].

2.35 In addition, some content that is not sexually explicit can encourage or normalise sexual violence. This can include, for example, threats of sexual assault or victim-blaming content which degrades survivors and victims of sexual violence.

2.36 Content that normalises sexual violence can also be created and shared via generative artificial intelligence ('GenAI') chatbots, which can replicate biased algorithms and training data.[67] [68] In the context of 'AI girlfriend' and 'AI boyfriend' chatbots, harmful gendered stereotypes about sexual relationships can be reinforced, for example, that women and girls should be submissive and always sexually available.[69]

2.37 We highlight good practice for addressing the circulation of content depicting, promoting or encouraging online gender-based harms, including sexual violence, in this Guidance.

Pile-ons and coordinated harassment

2.38 Within this section on pile-ons and coordinated harassment, we focus on misogynistic abuse that targets individual women and girls in public life. This covers the following kinds of illegal content and activity, addressed in the Illegal Content Judgements Guidance: harassment, threats and abuse and hate. It also covers the following kinds of content harmful to children, addressed in our Guidance on Content Harmful to Children: abuse and hate content and violent content.

2.39 Pile-ons and coordinated harassment are a common way that women and girls are targeted by harm online. Pile-ons and coordinated harassment involve many perpetrators targeting an individual victim or small group of victims with abusive, hateful or threatening content, often repeatedly or at scale.[70] This content can include demeaning attacks on the basis of gender or insulting and intimidating remarks.

2.40 It is important to note that critical or potentially offensive speech and expression is protected by human rights law, and is vital to maintaining a free and democratic society.[71] For example, people have the right to criticise women politicians or other women in public life because they disagree with their political actions or views, or their policies, and may do so in a way that shocks or offends. However, the right to freedom of expression is not absolute. Expression that promotes or justifies violence, hatred, xenophobia or another form of intolerance is not normally protected.[72] Illegal harassment, threats or abuse impact on other people's rights, and Parliament has determined that these offences give rise to priority illegal content under the Act.

---

[67] Feine, J., Gnewuch, U., Morana, S. and Maedche, A., 2020. Gender bias in chatbot design, *Chatbot Research and Design*. [accessed 31 October 2025].

[68] The Act applies to certain types of GenAI content, chatbots and services. See Ofcom's open letter to online service providers which outlines how the UK's Online Safety Act will apply to Generative AI and chatbots.

[69] Ofcom / Men and Boys Roundtable, 29 May 2025.

[70] Demos (Judson, E.), 2021. Silence, Woman: An investigation into gendered attacks online. [accessed 29 October 2025].

[71] See for example *Handyside v UK*: "Freedom of expression constitutes one of the essential foundations of such a [democratic] society, one of the basic conditions for its progress and for the development of every man. …. it is applicable not only to 'information' or 'ideas' that are favourably received or regarded as inoffensive or as a matter of indifference, but also to those that offend, shock or disturb the State or any sector of the population. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no 'democratic society'".

[72] Perinçek v. Switzerland [GC], 2015, § 230; Zemmour v. France, 2022, § 49; European Court of Human Rights, Key Theme – Article 10 Hate Speech.

2.41 Pile-ons and coordinated harassment can take many forms and the evidence on this harm area differs depending on what categories of harm are included and how they are defined.[73] Pile-ons and coordinated harassment can be perpetrated against men and boys, as well as women and girls. However, research demonstrates that pile-ons and coordinated harassment have disproportionate effect on women in public life,[74] and the kinds of content they are targeted with are often misogynistic and sexually violent.[75] These experiences can have a chilling effect on women and girls, whereby individuals suppress self-expression due to fears of legal penalties or social backlash, even without direct threats.

2.42 Women and girls in public life such as celebrities, politicians, journalists, influencers, athletes, and activists face heightened risks of pile-ons and coordinated harassment.[76] They are often targeted by organised gendered disinformation campaigns which weaponise false narratives to achieve social, political or economic aims.[77]

2.43 Ofcom research found that women in politics experience online hate and abuse consistently, often with a direct undertone of misogyny, and regularly including threats. The participants in this research told us that the hate and abuse is getting worse, becoming more common, sophisticated and seemingly normalised. The impact of these experiences has left women politicians feeling isolated, questioning their own abilities and adopting specific tactics to stay safe.[78]

2.44 It is not only women in politics who face a heightened risk of pile-ons and coordinated harassment. Research shows that women football players are more likely to receive online abuse than men football players,[79] and this abuse is often misogynistic in nature.[80] Similarly,

---

[73] 12% of UK women and 13% of UK men internet users aged 18+ experienced or saw "group shaming, boycotting, or excluding someone based on their views, opinions or actions, including online 'pile-ons'". This definition does not quite capture all the elements of pile-ons and coordinated harassment outlined in this Chapter. Source: Ofcom, 2025. Online Experiences Tracker – Wave 8.

[74] Research in 2023 from Fawcett Society found that 93% of the women MPs surveyed said that online abuse and harassment has a negative impact on how they feel about being an MP, compared with 76% of men. Source: Fawcett Society, 2023. A House for Everyone: A case for modernising Parliament. [accessed 25 September 2025].

[75] HM Government Stabilisation Unit, 2020. Quick-read guide: gender and countering disinformation. [accessed 29 October 2025]; US Department of State, 2023. Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors. [accessed 29 October 2025]. Research also shows for example that women scientists working on climate change were more likely to be targeted online because of their gender than men scientists. Source: Global Witness, 2023. Global Hating: How Online Abuse of Climate Scientists Harms Climate Action. [accessed 25 September 2025].

[76] HM Government Stabilisation Unit, 2020. Quick-read guide: gender and countering disinformation. [accessed 29 October 2025]; US Department of State, 2023. Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors. [accessed 29 October 2025].

[77] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S. and Kaufmann, Z.), 2021. Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online. [accessed 29 October 2025].

[78] Ofcom, 2025. Experiences of online hate and abuse among women in politics.

[79] This research was conducted during the 2023 Women's World Cup – women footballers were 29% more likely to be targeted with online abuse compared to the 2022 FIFA World Cup. Source: FIFA and FIFPRO, 2023. Social Media Protection Service Report: 2023 Women's World Cup. [accessed 21 October 2025]; Williams, A., Fielding-Lloyd, B., Newman, J. and Deller, R., 2025. Sports Organisations' Responses to Social Media Abuse Against Professional Sportswomen in UK Team Sports, *Communication & Sport,* 0 (0). [accessed 26 September 2025].

[80] Fenton, A., Ahmed, W., Hardey, M., Boardman, R., and Kavanagh, E., 2023. Women's football subculture of misogyny: the escalation to online gender-based violence, *European Sport Management Quarterly,* 24 (6). [accessed 26 September 2025]. See also: Ofcom and Kick It Out, 2025. Online hate and abuse in sport.

research on reality TV contestants found that gendered abuse online, often in the form of misogynistic and objectifying comments, disproportionately targets women.[81]

2.45 Research from the International Centre for Journalists found that nearly three quarters (73%) of women journalists had experienced online violence in some form in the course of their work. Journalists can receive thousands of abusive posts, often in response to reporting on politics or elections, as well as gender. Posts discredit their work and demean them personally. Often, they are called "stupid" or "hysterical" or their personal lives are investigated and criticised. One in four report receiving threats of physical violence.[82]

2.46 Pile-ons and coordinated harassment directed at women in public life aim to silence the individuals they target and also have a chilling effect on women and girls' participation more generally. Research from the Alan Turing Institute found that 77% of women are not comfortable expressing political opinions online because they fear being targeted with abuse[83] and a survey from Girlguiding found 36% of young women and girls were put off pursuing careers in sectors such as politics because of the abuse high-profile women face online.[84] Ofcom research found women politicians were deeply concerned about how online hate and abuse is dissuading women from entering politics, and what this might mean for the future of politics, the impact on democracy and the country as a whole.[85]

2.47 In the context of gender-based harms, pile-ons and coordinated harassment often overlap with misogynistic abuse and sexual violence, involving threats, descriptions of rape[86] and de-humanising or demeaning comments.[87] This form of coordinated behaviour also often involves doxing, where perpetrators publicly disclose private information about the individual targeted, such as their location, phone number or email address.[88] This can increase the risk of offline violence.[89]

2.48 Intimate image abuse (which we discuss in more detail under 'image-based sexual abuse') often plays a role in the coordinated harassment of women in the public eye. Research from Security Hero shows that 94% of sexual deepfakes depict women working in the entertainment industry[90] and deepfake intimate image abuse is increasingly used to target women politicians.[91]

---

[81] Demos (Judson, E.), 2021. Silence, Woman: An investigation into gendered attacks online. [accessed 29 October 2025].
[82] International Center for Journalists (Posetti, J. and Shabbir, N.), 2022. The Chilling: A global study of online violence against women journalists. [accessed 1 October 2025].
[83] The Alan Turing Institute (Stevens, F., Enock, F., Sippy, T., Bright, J., Cross, M., Johansson, P., Wajcman, J. and Margetts, H.), 2024. Understanding gender differences in experiences and concerns surrounding online harms: A nationally representative survey of UK adults. [accessed 3 November 2025].
[84] Girlguiding, 2022. Girls' Attitudes Survey 2022. [accessed 3 November 2025].
[85] Ofcom, 2025. Experiences of online hate and abuse among women in politics.
[86] Demos (Judson, E.), 2021. Silence, Woman: An investigation into gendered attacks online. [accessed 29 October 2025].
[87] Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive. [accessed 29 October 2025].
[88] The Global Partnership, 2023. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 29 October 2025].
[89] International Center for Journalists (Posetti, J. and Shabbir, N.), 2022. The Chilling: A global study of online violence against women journalists. [accessed 1 October 2025].
[90] Security Hero, 2023. State of Deepfakes: Realities, Threats, and Impact. [accessed 26 September 2025].
[91] Channel 4 (Newman, C.), 2024. Exclusive: Top UK politicians victims of deepfake pornography. [accessed 10 September 2025].

2.49 The functionalities and business model of a service – such as reposts[92] and recommender systems[93] – can amplify the virality of abusive content, including abusive content that is part of a pile-on.

2.50 In this Guidance, we explore good practice steps providers of user-to-user services can take to provide women and girls with increased control over their online experiences, including the ability to take action if they are the target of a pile-on or coordinated harassment campaign.

Stalking and coercive control

2.51 Within this section on stalking and coercive and control we focus on abusive behaviours typically carried out by a single perpetrator against a single victim. This covers the following kinds of illegal content and activity, addressed in the Illegal Content Judgements Guidance: harassment and stalking and controlling or coercive behaviour.

2.52 Technology facilitates existing dynamics of power and control, including stalking and coercive control. These forms of abuse can now be immediate and constant and reach a broad social network with minimal effort, having a faster and greater impact on different spheres of the survivor and victim's life.[94]

2.53 Evidence shows that perpetrators of stalking are more likely to target women.[95] Perpetrators of coercive control are also more likely to target women.[96]

[92] Harassment can begin when content is shared with a different audience to the one it was originally shared with, including through re-posts. Source: Thrasher, S.W., 2017. Yes, there is a free speech crisis. But its victims are not white men. The Guardian, 5 June, cited in Gosse, C., Veletsianos, G., Hodson, J., Houlden, S., Dousay, T.A., Lowenthal, P.R. and Hall, N., 2020. The hidden costs of connectivity: nature and effects of scholars' online harassment, *Learning, media and technology,* 46 (3). [accessed 10 November 2025].

[93] Institute for Strategic Dialogue (Thomas, E. and Balint, K.), 2022. Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the 'Manosphere'. [accessed 10 November 2025].

[94] Fernet, M., Lapierre, A., Hébert, M. and Cousineau, M., 2019. A systematic review of literature on cyber intimate partner victimization in adolescent girls and women, *Computers in Human Behaviour,* 100. [accessed 4 November 2025].

[95] The Crime Survey for England and Wales 2024 found that women are more likely to be the victim of stalking (20.2% of women aged 16+ experienced stalking at some point since the age of 16 compared to 8.7% of men). Women are also more likely to experience cyber-stalking (7.1% of women aged 16+ have experienced cyber-stalking at some point since the age of 16 compared to 3.7% of men). Source: Office for National Statistics (ONS), Crime Survey for England and Wales 2024. [accessed 19 September 2025].

[96] Evidence into coercive control perpetrated online specifically is limited but there is evidence that domestic abuse affects more women than men. The Crime Survey for England and Wales 2024 estimates that 9.5% of women and 6.5% of men experienced domestic abuse in the last year (2023-2024) and that 30.3% of women and 21.7% of men have experienced domestic abuse since the age of 16. Source: Office for National Statistics (ONS), 2025. Redevelopment of domestic abuse statistics: research update May 2025. [accessed 19 September 2025].

Stalking

2.54 Stalking is a form of harassment,[97] characterised by a pattern of fixated, obsessive, unwanted and repeated behaviour which is intrusive.[98] The majority of stalking is perpetrated by a partner or former partner. However, research reveals that perpetrators can include family members, friends, colleagues and strangers.[99]

2.55 Stalking has significant negative effects on survivors and victims, who may develop depression, anxiety, post-traumatic stress disorder and hypervigilance as a result of their experiences. A research study found that 87% of survivors and victims of stalking reported changes to at least one aspect of their life, including relationships, work and finances.[100]

2.56 Stalking frequently co-occurs across both online and offline spaces. Research from the Suzy Lamplugh Trust shows that 75% of stalking cases involved both online and offline behaviours.[101] Stalking presents a significant risk of escalation, with a research study finding stalking present in 76% of intimate partner homicides.[102]

2.57 Stalking can include a range of behaviour, from sending the victim persistent abusive messages to more covert forms such as monitoring their accounts or identifying their location from photos, which can enable offline stalking.[103] Perpetrators of stalking will often engage in repeated behaviours and use a variety of tactics to cause fear, alarm and distress to the victim.[104]

2.58 Stalking can be difficult for service providers to identify, as unlike some other harms it consists of a pattern of behaviour rather than a single piece of content. However, the risk of offline escalation presented by this harm means it is important for providers to take timely and proactive action when stalking is perpetrated on their service.

2.59 In this Guidance, we outline how providers can take proactive action against perpetrators of stalking, including allowing survivors and victims to report offline incidents of abuse, such as stalking, to the provider.

---

[97] Harassment is a pattern of oppressive and unreasonable behaviour carried out to cause the victim alarm or distress.

[98] More information about stalking is set out in Section 3 of the Illegal Content Judgements Guidance and Section 4 of the Illegal Harms Register of Risks.

[99] Suzy Lamplugh Trust, 2021. Unmasking Stalking: A Changing Landscape. [accessed 11 August 2025].

[100] Short, E. and Maple, C., 2011. The impact of cyberstalking: review and analysis of the ECHO pilot project. *Proceedings of the IADIS International Conferences – Web Based Communities and Social Media*. [accessed 29 August 2025].

[101] Suzy Lamplugh Trust, 2021. Unmasking Stalking: A Changing Landscape. [accessed 11 August 2025].

[102] This research was conducted in 10 US cities and focused on offline stalking behaviours. McFarlane, J., Campbell, J.C., Wilt, S., Ulrich, Y. and Xu, X., 1999. Stalking and Intimate Partner Femicide. *Homicide Studies,* 3 (4). [accessed 29 August 2025].

[103] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T. and Dell, N., 2018. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* [accessed 28 October 2025].

[104] Suzy Lamplugh Trust, 2021. Unmasking Stalking: A Changing Landscape. [accessed 11 August 2025].

Coercive control

2.60 Coercive control refers to the repeated or continuous perpetration of behaviour that is controlling or coercive, in the context of an intimate or family relationship.[105] Perpetrators of coercive control may also target the children of survivors and victims.[106]

2.61 Coercive control often involves different abuse types, co-occurring across multiple devices and services over many years. This includes abuse types that take place offline including physical abuse, or over different types of communications, such as financial and device-based abuse. Considering these harms in isolation can underestimate the scale and impact of abuse.

2.62 Coercive control is heavily under-reported. Half of survivors responding to a survey by Refuge said they told no one about the abuse and only a small proportion of survivors (13%) reported the abuse to the social media platform they experienced the abuse on. Only one in ten survivors felt empowered to report to the police.[107] Under-reporting happens due to shame about the abuse and lack of trust in service providers or the police to address the problem.[108]

2.63 Coercive control is a pattern of abuse that can include a range of behaviours, such as:

a) Device and app control: unauthorised access to a person's online accounts, by guessing passwords, or hacking.[109]
b) Monitoring and surveillance: monitoring messages and posts.
c) Impersonation, including catfishing: assuming the identity of a person to access private information; exploit, embarrass, discredit or shame them; contact or mislead them; or create fraudulent documents. This can happen via account hacking and can also include the online theft of documents.[110] This can extend to business profiles run by the survivor and victim to perpetrate economic harm.
d) Doxing: non-consensual public disclosure of private information, such as a victim's location, phone number or email address.[111] This can be particularly harmful in the context of honour-based violence, where leaked information or the threat of leaked information can lead to additional shame, stigma and violence.[112]

---

[105] More information about coercive and controlling behaviour is set out in Section 5 of our Illegal Harms Register of Risks.
[106] NSPCC Learning, 2023. The impact of coercive control on children and young people. [accessed 3 November 2025].
[107] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025].
[108] Flynn, A., Powell, A., Scott, A. and Cama, E., 2022. Deepfakes and Digitally Altered Imagery Abuse: A Cross-Country Exploration of an Emerging form of Image-Based Sexual Abuse. *The British Journal of Criminology*, 62 (6). [accessed 4 November 2025]; Ofcom Stakeholder Workshop 1 on Women and Girls Online Safety, 16 September 2024.
[109] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T. and Dell, N., 2018. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology. *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* [accessed 28 October 2025].
[110] Henry, N., Vasil, S., Flynn, A., Kellard, K. and Mortreux, C., 2021. Technology-Facilitated Domestic Violence Against Immigrant and Refugee Women: A Qualitative Study. *Journal of Interpersonal Violence*, 37. [accessed 29 October 2025].
[111] The Global Partnership, 2023. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 29 October 2025].
[112] End Violence Against Women Coalition and Faith and VAWG Coalition, 2021. Response to the Law Commission Consultation on Intimate Image Abuse. [accessed 29 October 2025].

e) Psychological or emotional abuse or threats: harassment and stalking, including sexual and/or dating harassment behaviours. This can include:

   i) Unwanted contact: sending the victim repeated messages.
   ii) Gaslighting: making the victim question their feelings and experiences.

2.64 Coercive control can be particularly hard for providers to identify. For example, a picture of a front door may seem innocuous to an outside observer or content moderator. However, for a victim who is fleeing an abusive partner and moving to a safe, secret location, receiving such a picture could be a sign that the user knows where they are. This can be traumatic for victims, and lead to them feeling physically unsafe because the perpetrator knows their location.[113] Furthermore, perpetrators can psychologically manipulate the victim to sow self-doubt and confusion in their mind, to conceal the abuse.

2.65 Issues with reporting and identifying coercive control point to the importance of qualitative work and the need for service providers to engage with support services to understand the nuances of this harm.

2.66 In this Guidance, we outline several good practice steps looking at how service providers can work with specialist services to support survivors and victims of stalking and coercive control. This includes consulting with subject-matter experts when setting terms of service and signposting users to supportive information.

Image-based sexual abuse

2.67 Within this section on image-based sexual abuse we focus on forms of abuse where the perpetrator uses intimate or indecent images to exert power and control over the victim. This covers the following kinds of illegal content and activity, addressed in the Illegal Content Judgements Guidance: intimate image abuse, self-generated indecent imagery, and cyberflashing.

2.68 Image-based sexual abuse refers to sharing or threatening to share intimate images depicting adults without consent[114] and the distribution, possession and publication of child sexual abuse material created by the child depicted in the image.

2.69 Image-based sexual abuse can form part of a pattern of harmful behaviour. It is not simply a product of online spaces but is also a manifestation of existing structures of sexual violence and misogyny.

2.70 Evidence suggests both intimate image abuse and cyberflashing disproportionately affect women and self-generated indecent imagery disproportionately affects girls. The Revenge Porn Helpline notes that 98% of the intimate images they have reported were of women[115] and research from Security Hero shows that 99% of deepfake intimate image abuse depicts women.[116] There are some types of image-based sexual abuse that disproportionately affect men, such as sextortion, with the Revenge Porn Helpline reporting that in 89% of

[113] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025].
[114] McGlynn, C. and Rackley, E., 2017. Image-Based Sexual Abuse, *Oxford Journal of Legal Studies*, 37 (3). [accessed 4 November 2025].
[115] South West Grid for Learning (Wright, D. and Mortimer, S.), 2024. Revenge Porn Helpline: Evidence to Women & Equalities Committee. [accessed 22 September 2025].
[116] This research was conducted from 15 July to 29 August 2023. Source: Security Hero, 2023. State of Deepfakes: Realities, Threats, and Impact. [accessed 26 September 2025].

sextortion cases the victim was a man.[117] Cyberflashing disproportionately affects women, who are more likely than men to be the targets of cyberflashing.[118] Self-generated indecent imagery disproportionately affects girls, with evidence showing that 98% of self-generated CSAM reports received by the Internet Watch Foundation in 2023 related to girls.[119]

Intimate image abuse

2.71 Intimate image abuse refers to sharing or threatening to share intimate images without the consent of the person shown in the image.[120] Creating[121] and taking[122] of intimate images without consent is also to be made illegal.

2.72 Threats to share intimate images and non-consensual sharing of intimate images can be perpetrated as part of a pattern of coercive control both online and offline.[123] Perpetrators may use threats to share intimate images to force survivors and victims to continue a relationship with them or allow the perpetrator to have contact with their children.[124]

2.73 Intimate image abuse can be perpetrated for financial, or other, gain (often known as 'sextortion'). Sextortion can be perpetrated by individuals but can also be carried out by organised criminal groups.

2.74 Non-consensual sharing of intimate images can also be perpetrated as part of a 'collector culture' where perpetrators exchange and discuss intimate images, often in communities dedicated to misogynistic abuse and sexual violence.[125] In this context, intimate image abuse can overlap with doxing, as images are shared alongside personal information identifying the survivor and victim.[126] These communities can develop on a range of services, including sites which are dedicated to hosting intimate image abuse content.[127]

2.75 Intimate images shared without consent may then be re-shared. Perpetrators often re-share images across multiple services, and in some cases images may go viral. This re-sharing means the images continue to circulate, causing re-victimisation and re-traumatisation for survivors and victims.

---

[117] South West Grid for Learning (Papachristou, K.), 2025. Revenge Porn Helpline 2024 Report. [accessed 27 August 2025].

[118] 20% of women aged 18 had received an unsolicited sexual photo compared to 11% of men aged 18+. Research was conducted in Great Britain. Source: YouGov (Smith, M.), 2024. More than a third of women under 40 have received unsolicited sexual photos. [accessed 30 September 2025].

[119] Internet Watch Foundation, 2023. IWF Annual Report 2023. [accessed 31 October 2025].

[120] Most commonly, an 'intimate image' is a photograph or video where the person or people are depicted engaging in, participating in, or are present during a sexual act and/or where their genitals, buttocks or breasts are exposed or covered only with underwear. An 'intimate image' also covers a photograph or video where the person or people are depicted in an act of, or carrying out personal care associated with, urination, defecation or genital or anal discharge. More information about intimate image abuse is set out in Section 6 of the Illegal Harms Register of Risks.

[121] See sections 66E – 66H of the Sexual Offences Act 2003. The provisions have not yet been commenced.

[122] Ministry of Justice and Alex Davies-Jones MP, 2025. Government cracks down on explicit deepfakes. [accessed 13 October 2025].

[123] Refuge, 2020. The Naked Threat. [accessed 30 September 2025].

[124] Refuge, 2020. The Naked Threat. [accessed 30 September 2025].

[125] Moore, A., 2022. 'I have moments of shame I can't control': the lives ruined by explicit 'collector culture', The Guardian, 6 January. [accessed 28 October 2025].

[126] South West Grid for Learning (Ward, Z.), 2021. Intimate image abuse, an evolving landscape. [accessed 30 September 2025].

[127] Henry, N., Flynn, A., 2019. Image-Based Sexual Abuse: Online Distribution Channels and Illicit Communities of Support, *Violence against women,* 25 (16). [accessed 30 September 2025].

2.76 Intimate image abuse has a profound negative impact on survivors and victims, including psychological, emotional, professional and relational harm.[128] For many survivors and victims of this kind of abuse, the impacts are devastating and all-encompassing, representing a "social rupture" that changes their lives irrevocably from that point onwards.[129] In some cases, survivors and victims can be coerced and exploited through the creation and circulation of intimate or sexual content, including through livestreams.

2.77 Intimate image abuse will also affect women in different situations differently. For example, sex workers' and adult content creators' experiences of intimate image abuse are often missed in safety interventions, and sex workers can be subjected to victim blaming and stigma. Sex workers are often targeted with blackmail, threats, doxing and stalking.130

2.78 Intimate image abuse can include both images that were taken or made consensually and images that were taken or made without consent. This includes images which have been artificially generated or manipulated, including deepfakes.131 Deepfake intimate image abuse often co-occurs with other online gender-based harms. For instance, it can be used by perpetrators to intimidate and discredit women in public life during pile-ons.

2.79 Deepfake intimate image abuse is an exponentially growing harm, with more deepfake intimate image abuse posted online in 2023 than in every previous year combined.[132] One reason for this growth in deepfakes is the availability of new tools powered by GenAI, which makes it easier for perpetrators to create realistic and life-like deepfake content, and a wider 'deepfake economy'. This includes widely available apps and sites that offer 'nudification' tools, forums where perpetrators provide instructions on how to create deepfake intimate image abuse content and websites that are dedicated to hosting this content.[133] The user base of these sites has dramatically increased – one reportedly receives 17 million views monthly – and they are easily accessed through internet search engines and advertised on social media sites, discussion threads and forums.[134]

2.80 In this Guidance, we include several good practice steps which could help service providers tackle intimate image abuse including steps focused specifically on preventing deepfake intimate image abuse. For example, we include introducing prompt and output filters to block a model from generating certain types of content which can include intimate images.

---

[128] Flynn, A., Powell, A., Scott, A. and Cama, E., 2022. Deepfakes and Digitally Altered Imagery Abuse: A CrossCountry Exploration of an Emerging form of Image-Based Sexual Abuse, *The British Journal of Criminology*, 62 (6). [accessed 4 November 2025].
[129] McGlynn, C., Johnson, K., Rackley, E., Henry, N., Gavey, N., Flynn, A. and Powell, A., 2021. 'It's Torture for the Soul': The Harms of Image-Based Sexual Abuse, *Social & Legal Studies*, 30 (4). [accessed 29 October 2025].
[130] Sanders, T., Trueman, G., Worthington, K. and Keighley, R., 2023. Non-consensual sharing of images: Commercial content creators, sexual content creation platforms and the lack of protection, *New Media & Society*, 27 (1). [accessed 28 October 2025].
[131] Ofcom, 2024. Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes.
[132] My Image My Choice, 2024. Deepfake Abuse: Landscape Analysis 2023-24. [accessed 28 October 2025].
[133] Ofcom, 2024. Deepfake Defences: Mitigating the Harms of Deceptive Deepfakes.
[134] Tenbarge, K., 2023, Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy, NBC, 27 March. [accessed 28 October 2025].

Self-generated indecent imagery

2.81 Self-generated indecent imagery (SGII) refers to child sexual abuse material created by the child depicted in the image.[135] [136]

2.82 SGII is often categorised into three types: consensual, non-consensual, and aggravated. These types refer to: images taken and shared consensually, images taken consensually that are shared onwards non-consensually, and images aggravated as a result of online grooming or financial coerced exploitation, respectively.[137] All forms of SGII are child sexual abuse material and therefore illegal content.

2.83 Regardless of the type, SGII can have negative mental health impacts. When sharing images amongst peers, the social repercussions that children may face from either consensual or non-consensual sharing of sexual images can be vast. This includes bullying, threats to share the images further and ostracisation from peer groups.[138]

2.84 The good practice steps set out in this Guidance are not aimed at addressing SGII. We have produced separate code measures focused on CSEA, including SGII. However, we do include steps to tackle intimate image abuse. Given the overlap in the contexts of how these images are often created and shared, interventions targeted at intimate image abuse, such as prompts, alerts, nudges and/or time-outsaimed at interrupting actions and encouraging reflection, may provide friction points for SGII.

Cyberflashing

2.85 Cyberflashing refers to the sending of an image of genitals for the purposes of causing alarm, distress or humiliation or for the purpose of obtaining sexual gratification.[139] [140] The term cyberflashing can also be used more generally to refer to the sending of sexual images without the consent of the recipient.

2.86 Cyberflashing breaches the privacy and sexual autonomy of the receiver of the image and can have negative psychological impacts on those targeted. Many survivors and victims describe the experience as aggressive or intimidating and feel frightened or vulnerable as a result.[141] Evidence suggests that women in minority ethnic groups and LGBTQ+ groups are

---

[135] We recognise the challenges associated with this terminology and how it may fail to capture the nature of the abuse suffered and unintentionally imply that a child is responsible for their own abuse. However, in the absence of a more appropriate and widely adopted alternative, we have chosen to use this wording to ensure clarity.

[136] More information about self-generated indecent imagery is set out in Section 2 of the Illegal Harms Register of Risks.

[137] INHOPE, 2025. What is Self-Generated CSAM? [accessed 4 September 2025].

[138] Schmidt, F., Varese, F., Larkin, A. and Bucci, S., 2023. The Mental Health and Social Implications of Nonconsensual Sharing of Intimate Images on Youth: A Systematic Review, *Trauma, Violence, & Abuse*, 25 (3). [accessed 4 September 2025].

[139] The cyberflashing offence refers to the sending or giving of a photograph or film of the genitals with the intent of causing alarm, distress, or humiliation, or for the purpose of sexual gratification on behalf of the sender (with recklessness as to whether alarm, distress or humiliation could be caused). See section 66A of the Sexual Offences Act 2003.

[140] More information about cyberflashing is set out in Section 19 of the Illegal Harms Register of Risks.

[141] Law Commission, 2022. Intimate image abuse: a final report. [accessed 28 October 2025].

disproportionately targeted by cyberflashing.[142] Cyberflashing is part of a wider harm in which women, particularly women in minoritised groups, are sexualised without consent.

2.87 Evidence suggests that services with particular functionalities are at higher risk of being used to perpetrate cyberflashing. These include direct messaging,[143] ephemeral messaging[144] and the ability to make connections with unknown users.[145]

2.88 In this Guidance, we set out good practice steps with particular relevance to providers of these services, including automated blurring of nudity content, with the option for users to unblur to signal their consent.

[142] McGlynn, C., 2021. Written evidence submitted by Professor Clare McGlynn, Durham Law School, Durham University. [accessed 29 October 2025]; McGlynn, C. and Johnson, K., 2022. Cyberflashing: Recognising Harms, Reforming Laws. [accessed 29 October 2025].

[143] YouGov (Smith, M.), 2018. Four in ten young women have been sent unsolicited sexual images. [accessed 24 September 2025].

[144] Ipsos UK, 2024. Online communications: Qualitative research exploring experiences of sexualised messages online among children. [accessed 10 November 2025].

[145] Ringrose, J., Regehr, K. and Whitehead, S., 2021. Teen Girl's Experiences Negotiating the Ubiquitous Dick Pic: Sexual Double Standards and the Normalization of Image Based Sexual Harassment. *Sex Roles,* 85 (558). [accessed 24 September 2025].

3. Taking responsibility

Overview

Context

3.1 Effective mitigation of online gender-based harms requires service providers to introduce appropriate governance and accountability processes, including compliance with the risk assessment duties set out in the Act.[146] This will ensure women and girls' online safety is a key consideration in service providers' design choices.

3.2 In this chapter, we set out actions service providers can build into their governance and risk assessment processes to appropriately address online gender-based harms. The specific actions include:

- **Action 1**: Ensure governance and accountability processes address women and girls' online safety.
- **Action 2**: Conduct risk assessments that focus on harms to women and girls.
- **Action 3**: Be transparent about women and girls' online safety.

3.3 For each action, we set out our expectation of what a baseline of safety looks like (**'foundational steps'**) for service providers to meet their duties to protect UK users. We have a range of enforcement powers to hold companies to account where they fail to comply with their duties.

3.4 We also highlight additional **good practice steps** to illustrate how providers can build on the foundational steps to prioritise women and girls' online safety, understand the risks and promote trust with their users.

3.5 We consider the application of these steps to be an ongoing exercise whereby providers continually re-assess and improve the experiences of women and girls on their service.

Our target outcomes

3.6 We expect that taking a safety-by-design approach to women and girls' online safety will ensure service providers consider online gender-based harms throughout the development lifecycle of their services.[147]

3.7 We also expect this will positively impact the online experience of all users. Specifically, we consider that marginalised and vulnerable users who experience disproportionate rates of harms are likely to experience improved protections by ensuring user safety is central to decision making structures across the organisation.[148]

---

[146] For more information, see our Illegal Content Risk Assessment Guidance, and our Children's Risk Assessment Guidance.

[147] Strohmayer, A., Slupska, J., Bellini, R., Neff, G., Coventry, L., Hairston, A. and Dodge, A, 2021. Trust and Abusability Toolkit: Centering Safety in Human-Data Interactions. [accessed 4 November 2025]; Chatham House (Wilkinson, I., Hofstetter, J.S., Shires, J. and Yahaya, M.S.), 2024. The role of the private sector in combatting gendered cyber harms. [accessed 4 November 2025].

[148] Constanza-Chock, S., 2020. Design Justice. [accessed 4 November 2025].

3.8 Gender-sensitive governance and risk assessment processes can also help service providers to better understand and anticipate risks to users. This improved awareness will increase the likelihood of these risks being appropriately prioritised and mitigated, including being factored into strategic decision making. We also recommend that service providers should make the outcomes of their governance and risk assessments accessible to external actors to increase accountability.

3.9 These governance and organisational design processes should help service providers in preparing to deal with changes in the online landscape that may increase risks to users, including sudden spikes in illegal content and sensitive events. They could also assist in monitoring and reviewing the effectiveness of measures designed to reduce risk.

Action 1: Ensure governance and accountability processes address online gender-based harms

3.10 Effective governance and accountability processes provide the foundation for service providers to identify, manage, and review risks to their users. By embedding accountability, oversight, independence, transparency, and clarity of purpose into their operations, providers can create well-functioning governance and organisational design processes. This could lead providers to respond more effectively to online gender-based harms, for example through senior leaders setting it as a priority.

3.11 Active and representative leadership is critical for a service provider to effectively respond to online gender-based harms that exist on, or are facilitated by, its service.[149] Senior leaders should be engaged with online gender-based harms and build this culture with relevant staff and decision-makers. This could include hiring a more diverse workforce — particularly ensuring representation of women and marginalised groups in leadership roles — to bring lived experience into decision-making processes. A culture of inclusion with clear internal roles, remits and accountability structures can strengthen a service providers' ability to identify, prevent, and respond to online harms against women and girls.

3.12 Governance and accountability measures can also support organisations prioritising safety-by-design, including developing internal (for example in learning and development goals) and external policies (for example in terms of service).

3.13 Service providers can decide which policies will be most appropriate for their service. However, in this Guidance, we outline recommendations and evidence on good practices that service providers can consider adopting to effectively address gender-based harms.

[149] Carnegie UK, The End Violence Against Women Coalition, Glitch, NSPCC, Refuge, 5Rights, Woods, L. and McGlynn, C., 2022. Violence Against Women and Girls (VAWG) Code of Practice. [accessed 4 November 2025]; Department for Science, Innovation & Technology and PUBLIC, 2025. Platform Design and the Risk of Online Violence Against Women and Girls. [accessed 29 October 2025].

Foundational steps: What are the expectations for service providers?

3.14 Our Codes of Practice set out the following steps for service providers, based on functionality, risk and size:[150]

a) **Board review**: Service providers' senior governance body, in relation to the service, should conduct and record an annual review of risk management activities pertaining to online safety. This would include assessing outstanding risks after implementation of the Codes of Practice measures. The review should include how a service provider is monitoring and managing emerging risks on their services.[151]

b) **Accountable individual:** Service providers should name an individual accountable to the most senior governance body for compliance with online safety duties.[152]

c) **Written statements of roles and responsibilities:** Service providers should have written statements that clearly show the roles and responsibilities for senior managers who make decisions about the management of online safety risks.[153]

d) **Internal monitoring and assurance function:** Service providers should establish internal systems to provide independent assurance that measures taken to mitigate and manage the risks of harm, as identified in the illegal harms and children's risk assessments, are effective and consistent.[154]

e) **Monitoring trends:** Service providers should track evidence of new kinds of illegal content, primary priority content that is harmful to children and priority content that is harmful to children on the service, and unusual increases in particular kinds of illegal content or content that is harmful to children, or proxies[155] for this. [156]

f) **Codes of Conduct:** Service providers should establish clear standards and expectations for individuals working for the provider around protecting users from online safety risks.[157]

---

[150] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this Guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).

[151] Illegal Content (ICU A1 / ICS A1), Protection of Children (PCU A1 / PCS A1).

[152] Illegal Content (ICU A2 / ICS A2), Protection of Children (PCU A2 / PCS A2).

[153] Illegal Content (ICU A3 / ICS A3), Protection of Children (PCU A3 / PCS A3).

[154] Illegal Content (ICU A4 / ICS A4), Protection of Children (PCU A4 / PCS A4).

[155] 'Proxy' means content that a provider determines to be in breach of its terms of service, where: a) the provider had reason to suspect that the content may be relevant [illegal content] [content that is harmful to children] [primary/priority content]; and b) the provider is satisfied that its terms of service prohibit the type of relevant [illegal content] [content that is harmful to children] [priority content] which it had reason to suspect existed.

[156] Illegal Content (ICU A5/ ICS A5), Protection of Children (PCU A5 / PCS A5).

[157] Illegal Content (ICU A6 / ICS A6), Protection of Children (PCU A6 / PCS A6).

g) **Terms of service and publicly available statements:** Service providers should publish clear and accessible provisions on how users are protected from: (1) illegal content, including illegal harms that disproportionately affect women and girls, such as stalking, coercive control and intimate image abuse; (2) content harmful to children including misogynistic abuse; and (3) non-designated content (see **Case study 1**).[158] The provisions should be easily accessible to users. This means they should be easy to find, clearly formatted, written to a comprehensible reading age for the youngest user permitted to use the service without parental consent, and designed to be compatible with assistive technologies like screen readers.

h) **Compliance training:** Service providers should ensure appropriate training for staff involved in the design and operational management of a service, with regards to their approach to compliance with online safety duties.[159]

i) **Duties about freedom of expression and privacy:** Service providers have a duty to have particular regard to the importance of protecting users' right to freedom of expression within the law, when deciding on, and implementing, safety measures and policies designed to ensure compliance with particular duties.[160] Similarly, service providers also have a duty to have particular regard to the importance of protecting users from a breach of any statutory provision or rule of law concerning privacy that is relevant to the use or operation of a user-to-user service (including concerning the processing of personal data).[161]

[158] Illegal Content (ICU G1 and G3 / ICS G1 and G3), Protection of Children (PCU G1 and G3 / PCS G1 and G3). Terms of service refer to user-to-user services, while publicly available statements refer to search services.
[159] Illegal Content (ICU A7 / ICS A7), Protection of Children (PCU A7 / PCS A7). In this context, staff refers to individuals working for the company.
[160] Section 22(2) (user-to-user services) and section 33(2) of the Act (search services). Under s.22(8) of the Act, these are safety measures and policies designed to secure compliance with any of the duties set out in section 10 (illegal content), section 12 (children's online safety), section 15 (user empowerment), section 20 (content reporting), or section 21 (complaints procedures). Section 33(4) of the Act provides the equivalent duties for search services: section 27 (illegal content), section 29 (children's online safety), section 31 (content reporting), or section 32 (complaints procedures).
[161] Section 22(3)(user-to-user services)and section 33(3) (search services). Under s.22 Category 1 services will have additional duties to undertake, publish and keep up to date impact assessments of both proposed and adopted safety measures and policies on users' right to freedom of expression within the law and also the privacy of users. Those services will also have a duty to specify in a publicly available statement the positive steps taken by the provider to protect users' right to freedom of expression within the law and privacy. These additional duties will come into force once the first register of Category 1 services is published.

3.15 **Case study 1** sets out examples of how a social media service provider can capture the risk of stalking in its terms of service and governance systems.

Case study 1 (foundational): Capturing stalking in terms of service

Scenario

A **social media provider** becomes aware through a campaign run by a frontline anti-**stalking** organisation that stalking perpetrators are exploiting its features, such as public posts, location data, and public lists of friends. When the service provider reviews its terms of service, the trust and safety team notices that the policy does not mention stalking.

Steps to take

The social media provider updates its **terms of service**, including its community guidelines, to define and prohibit stalking, as well as set out what the provider will do to act on breaches of this policy clearly and explicitly. The service provider's legal and trust and safety teams work together to ensure these terms of service are clear and accessible and define stalking precisely to ensure clarity of scope.[162]

- Once agreed, the service provider needs to update and publish the guidelines for their internal processes and systems. The new guidelines inform updated staff training for trust and safety, content moderation, and leadership teams.
- The service provider needs to ensure the published terms of services are: (1) drafted using clear and accessible language for the targeted audience; (2) written to a comprehensible reading age for the youngest user; (3) clearly formatted; and (4) easy to find and accessible to those using assistive technologies.
- After publication, the service provider establishes a process for updating its internal content policies in response to evidence of new and increasing illegal harm on the service (see Action 2).

Considerations

- Changes to service content rules are likely to amount to a "significant change" to the service, requiring a new risk assessment relating to the proposed change to be carried out prior to its implementation. The Illegal Harms and Children's Risk Assessment Guidance provides further guidance on what amounts to a "significant change" for these purposes and how the service provider should consider this.

[162] See Protection from Harassment Act 1997 (section 2A). Criminal Justice and Licensing (Scotland) Act 2010 - section 39.

Good practice steps: How can service providers go further?

3.16 Service providers can go further by demonstrating both internally and externally their commitment to women and girls' online safety and their willingness to improve in response to feedback.

3.17 In this section, we set out examples of good practice steps that service providers can take in tandem with the foundational steps to improve the protection of users against online gendered-based harms. This can include:

a) **Setting policies** that are designed to tackle specific forms of online gender-based harms that are prevalent on the service (see **Case study 2**).[163] This could include defining and prohibiting:

   i) Misogynistic abuse affecting specific groups, such as misogynoir (hate directed at Black women and girls)[164] and deliberate misgendering (referring to someone, especially a transgender person, using a word, especially a pronoun or form of address, which does not reflect their gender identity);[165] or
   ii) Promotion of offsite gender-based harms, such as promoting deepfake creation services like nudification apps, or forums sharing explanations of how to monitor your partner.

b) **Considering intersectionality of harms** in the governance and decision-making processes undertaken by service providers. For example, service providers should also consider how users' gender identity, race or disability may increase the risks they face. This understanding can help inform governance and decision-making processes by avoiding addressing harms in isolation.

c) **Consulting with subject-matter experts**, particularly those with experience of supporting survivors and victims of gender-based harms and those with expertise in understanding how different types of content and activity can lead to harm, when setting policies and terms of service and developing staff training (see **Case study 3**). Appropriate subject-matter experts will vary depending on the service provider. For instance, services that allow pornography could consult with an organisation that provides classifications, such as the British Board of Film Classification, when considering how to address potentially harmful kinds of sexually explicit content.

[163] Service providers should have regard to the needs of their UK user base in considering what languages are needed to ensure their policies are accessible.

[164] Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive. [accessed 29 October 2025]; Bailey, M., 2021. Misogynoir Transformed: Black Women's Digital Resistance. [accessed 4 November 2025].

[165] It is important to note that some gender critical beliefs are protected as "philosophical beliefs" under the Equality Act 2010 (For example, in *Forstater v CDG Europe* [2021] 6 WLUK 104, the Employment Appeal Tribunal held that a "gender-critical" belief that sex was biologically immutable, and that sex rather than gender identity was fundamentally important, was a "philosophical belief" protected under the Equality Act 2010). However, some forms of misgendering, as set out in the Children's Harms Guidance, are abusive and/or hateful and can constitute harassment.

d) **Training employees** responsible for and/or participating in setting policies and governance or decision-making processes on evolving online gender-based harms, including how safety-by-design can mitigate these harms. Training emphasises consistent, proportionate enforcement, and proactively addresses misogynistic abuse and sexual violence, pile-ons and coordinated harassment, stalking and coercive control, and image-based sexual abuse while not unduly censoring debate. It can also include information to ensure that contractors receive adequate training and support in line with the above.

e) **Ensuring adequate resourcing** to develop ongoing policy and risk expertise needed to deliver trust and safety objectives, such as setting policies or conducting a gender-sensitive risk assessment. This could also include adequate resource and expertise to respond to moments of increased risks of online gender-based harms, such as during periods of key civic moments such as elections or cultural events.[166]

f) **Creating a media literacy-by-design policy** to promote critical and informed use of its service, as set out in Ofcom's Best Practice Principles for Media Literacy by Design. Media literacy interventions can be particularly important and valuable for children and young adults.

g) **Establishing an oversight mechanism** for governance and design decisions, including risk assessments, trust and safety policies and their implementation across content moderation teams (see **Case study 4**).[167]

[166] Ofcom, 2025. Online hate and abuse in sport. Research has found women in public life experience disproportionate and unique harms, including during election periods. Institute for Strategic Dialogue, 2025. Crushing Comments: Gendered Harassment During the 2024 EU Parliament Elections on TikTok. [accessed 17 November 2025]; Institute for Strategic Dialogue, 2024. Votes and Vitriol: Online Abuse Targeting Women Candidates in the 2024 French Legislative Elections. [accessed 17 November 2025]. Ofcom, 2025, Experiences of online hate and abuse among women in politics.

[167] LEAF (Khoo, C.), 2021. Deplatforming Misogyny: Report on Platform Liability for Technology-Facilitated Gender-Based Violence. [accessed 4 November 2025].

3.18 **Case study 2** sets out examples of how a discussion forum can ensure its terms of service are updated to address new types of online gender-based harms.

Case study 2 (good practice): Clear gender-based harm policies

Scenario

A **discussion forum** notices that several forums have been set up to re-share images of women and repost them with objectifying and degrading comments based on gender.

The discussion forum already prohibits intimate image abuse and most forms of **misogynistic abuse and sexual violence,** but it realises the policy does not specifically address this type of misogynistic abuse.

Steps to take

- The discussion forum updates its misogynistic abuse and sexual violence policy to make it clear that targeting a woman or girl with objectifying or degrading comments is a violation of its **terms of service**.
- The discussion forum's legal and trust and safety teams work together to ensure the new policy is clear and accessible and defines this type of misogynistic abuse and sexual violence precisely to avoid overreach.
- Due to this type of behaviour being clearly prohibited in the discussion forum's policies, survivors and victims can report this specific harm through the appropriate and established reporting systems and processes (see Action 8).
- The service provider can subsequently use the reported incidences as an information source for evaluating the effectiveness of their policy and update the policy as or if needed.

Considerations

- Ultimately, it is up to providers to decide what content they allow on their services as long as this complies with the duties set out in the Act, but providers should be transparent with users about the choices they make and content that users may be exposed to.
- Developing and implementing policies is only effective if a service provider commits to their enforcement, including taking appropriate action for breaches of the policies (see Action 9).

3.19 **Case study 3** sets out examples of how a messaging service provider can engage with expert organisations to inform its policy development.

Case study 3 (good practice): Engaging with subject-matter experts

Scenario

A **messaging service provider** becomes aware that perpetrators of **coercive control** are using its service to intimidate and monitor their victims. These behaviours often manifest as patterns, making them difficult to detect in isolated instances like a single message or post. The messaging service struggles to anticipate how users may misuse its features, especially as abusive tactics evolve rapidly.

Steps to take

The service provider seeks out **organisations that support and represent survivors and victims** of coercive control, particularly organisations led by and for survivors and victims.

- The service provider assesses what expertise it needs and what organisations are best placed to provide appropriate and informed input. It meets with organisations that represent specific groups at risk of harm - such as groups representing LGBT+ or immigrant survivors and victims - to include intersectional experiences of abuse. Initial bilateral meetings lead the service provider to set up an advisory board which meets regularly for formal input and review.
- When rolling out new features, the service provider uses the method of co-design by bringing in survivors and victims through the advisory board earlier on in the design process as active participants, rather than consulting with them as an afterthought following the development of a feature. The service provider and organisations work closely together to ensure this engagement is sensitive and trauma informed.
- After engaging with the advisory board, the provider can better assess the risk of coercive behaviour on their service (see Action 2) and develop better policies (see Action 1), preventative safety measures (see Actions 4-6) and reporting systems (see Action 8) in response to this harm.

Considerations

- Engagement with providers can put a burden on vulnerable groups, as well as organisations that support them.[168] As these organisations are often under-resourced, it is good practice to provide appropriate compensation for any work dedicated to improving the provider's policies or practices (this includes external assessors for risk assessments).[169]
- Participatory methods like co-design can also be resource intensive. This may be cost prohibitive for smaller service providers. If so, providers could consider drawing on existing research by organisations with subject matter expertise in online gender-based harms, and organisations experienced in trauma-informed research (see **Case study 6**).

[168] Ofcom Stakeholder Workshop 2 on Women and Girls Online Safety, [accessed 19 November 2025].
[169] Women's Aid, 2024. Domestic abuse services struggling to fill critical gaps in a challenging landscape, exacerbated by the rising cost-of-living. [accessed 10 November 2025].

3.20 **Case study 4** sets out examples of how a social media provider can leverage external oversight for content moderation.

Case study 4 (good practice): External oversight for content moderation

Scenario

A **social media provider** is concerned after the misapplication of their content moderation policies has led to the removal of non-prohibited content, such as mothers breastfeeding, LGBTQ+ couples kissing or survivors and victims' sharing their experiences of sexual violence.[170]

The social media provider notes that the content moderation algorithms have biases due to their training data, which may contribute to the inconsistent moderation of content, alongside moderator training that may miss important considerations

Steps to take

The service provide sets up an **external arbitration process**, by involving a panel of subject matter experts to review moderation decisions:

- The arbitrator accepts complaints from users appealing content moderation decisions and chooses to review certain complaints that demonstrate inconsistent enforcement against the service provider's policies. Following this review, the arbitrator will recommend a decision is either overturned or upheld.
- The service provider uses feedback from the arbitrator and data from its appeals to identify patterns of bias in content moderation.[171] It uses this data to inform future reviews of its content moderation algorithms (see Action 6) and inform its training for content moderation teams (see **Case study 23**).

Considerations

- Setting up an arbitrator may be cost prohibitive for smaller or less established service providers. Smaller providers could consider drawing on existing research by organisations with subject matter expertise in online gender-based harms, and engage in user research to understand where moderation decisions have had unintended effects (see **Case study 6**).

[170] Peters, J., 2020. Sexual Content and Social Media Moderation, *Washburn Law Journal*, 59 (3). [accessed 4 November 2025]; Institute of Strategic Dialogue (Matlach, P-D. and Small, A.C.), 2024. Off-limits: Sexual Violence on TikTok. [accessed 4 November 2025].

[171] Hawkins, I., Roden. J., Attal, M. and Aqel, H., 2023. Race and gender intertwined: why intersecting identities matter for perceptions of incivility and content moderation on social media, *Journal of Communication*, 73 (6). [accessed 4 November 2025]

Action 2: Conduct risk assessments that focus on harms to women and girls

3.21 Evidence suggests that risks of online gender-based harms can be broadly overlooked and culturally diminished in organisations.[172] For example, gendered threats like intimate partner violence have often been missed in industry and research threat modelling processes which look at how a system's security might be compromised.[173] Instead, security evaluations can often assume that the main source of threat is an external or unknown stranger. As a result, there are insufficient safety features or processes to respond when the threat is an intimate partner of the user.

3.22 For these reasons, gender-sensitive risk assessments are important to effectively mitigate against such harmful online activity. Gender-sensitive risk assessments are achieved by making sure that existing risk assessment processes (as set out in our Illegal Content and Children's Risk Assessment Guidance) consider gender as part of the user base demographics. Services can further build on these existing risk assessments to capture the dynamics of gender-based harms.[174]

3.23 Service providers need to build an understanding of factors that enable and promote online gender-based harms in their online safety risk assessments to design safer systems and processes. As technologies evolve rapidly, providers will also need to keep track of emerging risks and trends in perpetration, particularly when rolling out new measures and features.

Foundational steps: What are the expectations for service providers?

3.24 Our Codes and risk assessment guidance set out the following steps for service providers, based on functionality, risk and size:[175]

a) **Risk assessment**: Service providers have a duty to conduct a suitable and sufficient illegal content risk assessment, and take appropriate steps to keep it up to date. They must also carry out a further risk assessment when they propose to make a significant change to the service, relating to the impacts of the proposed change.[176] Similarly,

[172] Criado-Perez, C., 2019. Invisible Women. [accessed 4 November 2025].
[173] Slupska, J. and Tanczer, L., 2021. Threat Modeling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things in *The Emerald International Handbook of Technology-Facilitated Violence and Abuse.* [accessed 4 November 2025].
[174] Enhanced processes for product testing, such as abusability and red teaming, can be particularly relevant in highlighting risks that disproportionately affect women and girls. These will be explored further in Chapter 4, with Chapter 3 focussing on overarching risk assessment and threat modelling practices that should take place before new products are developed.
[175] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this Guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[176] Section 9 (user-to-user) and section 26 (search) of the Act. All providers have to assess the risk of harm to all users from illegal content. In the context of illegal content that disproportionately affects women and girls, this includes controlling and coercive behaviour content, stalking, and intimate image abuse.

providers whose services are likely to be accessed by children have a duty to conduct suitable and sufficient children's risk assessments.[177] To note, some service providers will need to conduct additional assessments under the forthcoming user empowerment duties. Our guidance suggests four steps for a suitable and sufficient risk assessment:[178]

iii) **Understand the harms**: Service providers must assess the level of risk for each kind of priority illegal content and other illegal content, as well as for content harmful to children. This includes content and activity which disproportionately affects women and girls, such as intimate image abuse and exposure to pile-ons and coordinated harassment.

iv) **Assess the risk of harm:** Where applicable, service providers should consider the risk of harm with regards to their user base, business models, functionalities, governance, and systems and processes. We provide guidance on the evidence that service providers should use for this, identifying "core inputs" and "enhanced inputs". For example, we consider results of product testing to be an enhanced input that service providers can use to assess the risk of harm on their services (see Action 4). Service providers can also consider data or information from their reporting and complaints procedures to assess the prevalence of particular kinds of harms, and the service's track record in handling such complaints.[179]

v) **Decide measures, implement and record**: If service providers implement measures recommended in the Codes of Practice, they will be treated as complying with the relevant duties.The Act does not require that service providers adopt the measures set out in the Codes. Services can choose to comply with their duties in an alternative way that is proportionate to their circumstances as long as they keep a record of what they have done and how the relevant duties have been met. Service providers should record the outcomes of their risk assessment.[180]

vi) **Report, review and update risk assessments**: Service providers should report their risk assessment outcomes and online safety measures to a relevant internal governance body and update them regularly.

b) **Internal content and search moderation policies**: These policies should be set having regard to the findings of the risk assessment and providers should have processes in place for updating these policies in response to evidence of new and increasing illegal harm or harm to children on the service.[181]

---

[177] Section 11 (user-to-user) and section 28 (search) of the Act. Providers of sites accessible to children have to assess the risk of harm to children from content harmful to them, including violent, abusive, hateful and pornographic content.

[178] For more information, see our Illegal Content Risk Assessment Guidance, and our Children's Risk Assessment Guidance.

[179] For additional information on (1) identifying illegal content and content harmful to children, refer to the Illegal Content Judgements Guidance or Guidance on Content Harmful to Children; and (2) understanding how these harms manifest, refer to the Illegal Harms and Children's Registers of Risks.

[180] For more information, see our Record-Keeping and Review Guidance.

[181] Illegal Content (ICU C3 / ICS C2), Protection of Children (PCU C3 / PCS C3).

3.25 **Case study 5** sets out examples of how a gaming service can account for the risks of harm to different users, considering the overlap of harms and the influence of age on risk, in line with our risk assessment guidance.[182]

Case study 5 (foundational): Gender-sensitive risk assessment

Scenario

Following an investigation by a children's rights group, a **gaming service** becomes aware that girls and boys could be facing different types of risks on their services. The service provider previously conducted a risk assessment, in line with their statutory duties under the Act, but they did not consider the risks experienced by different genders in depth.

Steps to take

The service provider appropriately trains and resources its staff to **conduct a gender-sensitive risk assessment** as part of an update to its risk assessment for illegal content and content that is harmful to children.

- The service provider assesses its data on user base demographics to understand the risks experienced by different genders, including which harms disproportionately affect girls.[183]
- The service provider maps how both girls and boys with multiple protected characteristics experience unique and compounding risks in their risk assessments, taking account of Ofcom's risk profiles.[184]
- The service provider also learns that young women (aged 18-34) are particularly at risk of image-based sexual abuse including cyberflashing and intimate image abuse on messaging functionalities.[185]
- It also learns girls and young women from the most deprived areas, LGBTQ+ girls and disabled girls are more likely to see hate speech on its services.[186]
- As a result of this risk assessment, the service provider makes changes to its default settings to ensure the specific risks that girls and young women face are addressed, resulting in changes that mean all users are safer (see Action 5).

Considerations

- When considering the use of personal data, providers must also consider privacy rights and comply with data protection law requirements, including the data minimisation principle.[187]

[182] Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance.
[183] The ICO provides useful guidance on the use of storage and access technology, which will be relevant in cases where collection of demographic data is facilitated by these technologies. The ICO has broader guidance and resources on the Privacy and Electronic Communications Regulations (PECR) and the UK GDPR.
[184] Illegal Harms Register of Risks and Children's Register of Risks.
[185] For more information, see section 6 (intimate image abuse) and section 19 (cyberflashing) in the Illegal Harms Register of Risks.
[186] Girlguiding, 2024. Girls' Attitudes Survey 2024 [accessed 19 August 2025].
[187] We encourage providers to consult the ICO guidance on UK GDPR requirements and Anonymisation. Providers are also encouraged to consult the Age-Appropriate Design Code when processing the personal information of children. [accessed 4 November 2025].

Good practice steps: How can service providers go further?

3.26 Service providers can gain additional insights into how design choices create risk for women and girls by seeking expert advice and hearing directly from users about their experiences. These insights can also help service providers better understand how best to mitigate gender-based harms on their services.

3.27 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps to improve their risk assessments. This can include:

a) **Using external assessors** to monitor the threat landscape, including local partners with regional and cultural knowledge (e.g., civil society organisations or law enforcement), and international partners with expertise in highly contextual risk areas such as stalking and coercive control.[188] [189]

b) **Engaging with subject-matter experts and vulnerable groups**[190]**,** such as survivors and victims (including girls and young women), to better understand their experiences (see **Case study 3**).[191] [192]

c) **Conducting user research,** such as surveys, to better understand users' preferences and experiences of risk.[193] This should include capturing how risk manifests for different types of users (for example, identifying what is risky for a content creator compared to a content consumer).[194]

d) **Conducting additional assessments to assess impacts on users' self-expression, freedom from discrimination and privacy**,[195] [196] especially for those with protected

---

[188] Carnegie UK, The End Violence Against Women Coalition, Glitch, NSPCC, Refuge, 5Rights, Woods, L. and McGlynn, C., 2022. Violence Against Women and Girls (VAWG) Code of Practice. [accessed 4 November 2025]; Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive. [accessed 29 October 2025].

[189] This could be an example of an 'enhanced input' under the risk assessment process laid out in the Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance (namely seeking the views of independent experts). We expect enhanced inputs to be referred to by some kinds of services to ensure their illegal content and children's risk assessments are suitable and sufficient, but are optional for others, and good practice in this regard.

[190] When engaging with victims and survivors and users with protected characteristics, providers may collect special category and/or criminal offence data, both of which require additional legal protection under data protection law. For more information, please see the ICO's guidance on special category data and criminal offence data. This guidance is also relevant to good practice step below concerning user research.

[191] Ofcom Stakeholder Workshop 2 on Women and Girls Online Safety, 19 November 2024; Women's Aid, 2024. Domestic abuse services struggling to fill critical gaps in a challenging landscape, exacerbated by the rising cost-of-living. [accessed 10 November 2025].

[192] This could be an example of an 'enhanced input' under the risk assessment process laid out in the Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance (namely engaging with relevant representative groups). See further above.

[193] Service providers should have regard to the needs of their UK user base in considering what languages are needed to ensure their user surveys are accessible.

[194] This could be an example of an 'enhanced input' under the risk assessment process laid out in the Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance (namely consultation with users and user research). See further above.

[195] Impacts on users' privacy will often be considered under a data protection impact assessment (DPIA). A DPIA is a separate legal requirement under data protection law where services undertake processing of personal data that is likely to result in a high risk to the rights and freedoms of individuals. This includes certain specified types of processing, and the ICO has developed a screening checklist to help determine when a DPIA is necessary. For more information, see the ICO's guidance on Data Protection Impact Assessments (DPIAs).

[196] See footnote 162 for more information about the duty on providers of Category 1 services to carry out privacy and freedom of expression impact assessments.

characteristics.[197] For example, services can evaluate algorithmic systems such as content moderation and recommender systems for a variety of risks in relation to bias and discrimination.[198] Regular and periodic testing and evaluation of their systems (including pre- and post- deployment) can help services consider the risk their algorithmic systems pose, including bias and discrimination.[199]

e) **Considering additional supply chain risks** that may impact the effectiveness or quality of content moderation processes, systems and decisions**,** including risks from outsourcing and business relationships.[200] For example, third party contractors hired for content moderation may not have received best practice training in online gender-based harms (see Action 1).

[197] Equality and Human Rights Commission, 2019. Human Rights and Business. [accessed 16 October 2025]; United Nations Human Rights Office of the High Commissioner, 2011. UN Guiding Principles on Business and Human Rights.

[198] User-to-user services should also ensure that children's recommender feeds exclude or limit the prominence of content harmful to children, which may also involve undertaking algorithmic assessments. See the foundational steps under Action 6 and Case study 10 for more information.

[199] For further information on assessments of algorithmic systems please see: Metcalf, J. et al., 2021. Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts [accessed 20 August 2025].; Ashar, A. et al., 2024. View of Algorithmic Impact Assessments at Scale: Practitioners' Challenges and Needs. [accessed 20 August 2025].

[200] 5Rights Foundation (2025) Children & AI Design Code: A protocol for the development and use of AI systems that impact children, pp. 15-16; EVAW (2022) Violence Against Women and Girls (VAWG) Code of Practice.

3.28 **Case study 6** sets out examples of how a dating service provider could partner with a specialist organisation to complete trauma-informed research and user surveys to build its understanding of what harms are occurring on its services.

Case study 6 (good practice): Trauma-informed research and user surveys

Scenario

A **dating service provider** suspects that users are experiencing **stalking and coercive control**, however they do not have sufficient insight into when and how these harms are occurring on their service. To uncover these risks and understand user experiences, the service provider recognises the need to gather information directly from their users, which differs from its usual user experience testing, in a sensitive and effective manner.

Steps to take

The dating service provider **partners with a specialist organisation**, to design and deliver trauma-informed user surveys that are sensitive to the trauma experienced by survivors and victims in order to facilitate their engagement. [201]

- The specialist organisation encourages the service provider to apply trauma-informed design principles, such as providing clear information on consent and data use, prioritising user privacy, and localising surveys (for multi-country surveys) with teams trained in trauma awareness.
- To ensure users receive appropriate support, the survey results should signpost to services that offer supportive information, additional personalised support on sexual trauma and access to a trauma-informed therapist, depending on the users' specific results (i.e., the results indicated a need for additional support).
- The dating service provider then uses insights from the surveys to develop safety tools and monitor the experiences of users engaging with the service. This helps the provider identify trends or issues before they become widespread and evaluate safety measures (see Chapter 4 for further details on preventative measure).

Considerations

- In the survey design, the provider should be careful to avoid re-traumatisation and any infringements of data protection requirements.
- When collecting personal data from users, providers must also consider privacy rights and comply with data protection law requirements, including the data minimisation principle.[202]

[201] Chayn, 2024, How can we make quantitative research more trauma-informed? [accessed 14 October 2025]; Chayn, 2023, Chayn's trauma-informed design principles [accessed 14 October 2025].

[202] Whenever possible, providers should anonymise or pseudonymise personal information to reduce the risk of it being linked to an identifiable individual. We encourage providers to consult the ICO's Anonymisation Guidance. Personal information gathered directly from users to identify risks may constitute special category data and/or criminal offence data, both of which require additional legal protection under data protection law. For more information, see the ICO's guidance for services regarding special category data and criminal offence data.

Action 3: Be transparent about women and girls' online safety

3.29 Transparency is crucial to strengthen safety governance of online services to ensure user safety is considered in the design and operation of the service, provide users with more meaningful control over their online experiences, and promote trust in services' safety measures.

3.30 As laid out in our Final Transparency Guidance, the Act gives Ofcom powers to require providers of categorised services to publish certain information about their service in annual transparency reports based on requirements that Ofcom will issue to service providers via a transparency notice.

3.31 Separately, Ofcom must also produce and publish its own transparency report at least once a year summarising insights and conclusions drawn from the transparency reports produced by providers. This may include identified patterns and trends, good industry practice, and any additional information Ofcom considers relevant to help contextualise those findings for the public.

3.32 While providers' transparency reports will typically offer insight into how individual services address risks and mitigate harms on their services, Ofcom's transparency reports will contextualise those findings, drawing out points of comparison between services and highlighting examples of good and poor practice for the wider industry.

3.33 Our aim is for annual transparency reports to encourage services to improve their safety systems and processes. Publishing data on how providers of categorised services apply their policies and test the effectiveness of safety features or innovations will allow for a clearer assessment of what works, and also allows interventions to improve user safety or to deter online gender-based harms to be shared and adopted more widely.[203] We expect to use transparency reporting to equip stakeholders, including Ofcom and civil society, with more information about the safety practices of providers of categorised services and their effectiveness. This will aid stakeholders in understanding best practice and encourage evidence-based safety improvements at services.

3.34 Furthermore, increased information in the public domain may encourage providers to consider the impact and public opinion of the measures they have adopted (or not adopted) to tackle online gender-based harms. We intend to hold services to account by commenting on gaps in their systems and processes and enable learning from practices across the wider industry.

3.35 Our transparency reports will also empower UK users with relevant and accurate information about risks and safety outcomes on services so that they can take informed decisions about how to live their lives online. This could be particularly relevant for women and girls who face disproportionate risk online and therefore must enable them to make informed choices to foster safe online experiences.

[203] eSafety Commissioner, 2024. Technology, gendered violence and Safety by Design. [accessed 20 November 2025].

Foundational steps: What are the expectations for service providers?

3.36 Under the Act, there are specific duties pertaining to transparency that are foundational steps:

a) **Transparency reports:** Categorised services need to comply with additional duties, including a duty to publish transparency reports.[204]

3.37 Once a year, Ofcom must issue providers of categorised services with a transparency notice requiring them to produce a transparency report about their service.[205]

3.38 An important component of the decisions we make about our transparency notices will be what information will be reported consistently over time – what we are calling 'core' information – and what information will be reported on an ad hoc basis, within the context of a policy area that we have identified based on our regulatory focus each year – what we are calling 'thematic' information. For instance, for one reporting year we may seek to prioritise information about specific priority illegal harms that disproportionately affect women and girls.

3.39 The matters that we can request services to produce in transparency reports are listed in Schedule 8 to the Act which covers a wide variety of matters relating to online safety. In our Final Transparency Guidance, we set out illustrative examples of the information that we might request, such as the incidence of illegal content or content harmful to children.[206]

3.40 We plan to conduct engagement activities in each annual reporting cycle. For example, if our chosen thematic area relates to online gender-based harms, we may engage with relevant stakeholders with expertise in this area to inform the design of our transparency notices, or to gain understanding about how we can convert data into insights and make our reports useful and accessible for our audiences.

3.41 Following the publication of providers' reports, we will publish our own report to provide insights for users from the providers' reports. Such insights will highlight the implications of reported data about risks and safety outcomes, so that users can make informed choices about the services they use. Once published, we may look to engage with expert organisations (among others) to discuss our report and any recommended improvements for future reporting.[207]

Good practice steps: How can service providers go further?

3.42 Although the transparency reporting duties in the Act apply to categorised services, all services can improve accountability and better inform users by increasing transparency of their operations. For example, a service provider can improve transparency by producing voluntary transparency reports or publishing ad hoc information about emerging trends and updates relevant to online gender-based harms.

---

[204] See section 95 of the Act. Ofcom will publish a register of Category 1, 2A and 2B services and keep it up to date. Categorised services will need to comply with a series of additional duties, and for all categorised services this includes transparency reporting.
[205] See section 95 of the Act.
[206] Ofcom's Final Transparency Guidance, paragraph 3.4.
[207] Ofcom's Final Transparency Guidance, paragraph 5.5.

3.43 Increased transparency about their operations can enable providers to demonstrate to users the actions they take and their impact on women and girls' online safety.

3.44 The good practice steps providers could take include:

a) **Sharing information about the prevalence of different forms of online gender-based harms** where that data is already available to providers and collected and managed in line with data protection regulations. This information could include data on user reports and their outcomes (removals, timelines, and appeals), and, where possible, disaggregated data on outcomes, for example by user demographics. This should be in line with data protection law. It could provide the public – including researchers and civil society organisation – with additional information to understand which harms disproportionately affect women and girls, as well as indicate any biases in how reports are dealt with.[208] Further, it could include specific data on the experiences of marginalised groups, including, for example, disabled women.

b) **Sharing evidence on emerging trends and risks,** to the extent relevant and at an appropriate level of detail**,** related to online gender-based harms with key actors involved in prevention, such as civil society, law enforcement, and researchers.

c) **Sharing information about additional assessments undertaken (see Action 2) and the effectiveness of measures in place** to address online gender-based harms. For example, sharing the outcomes of algorithmic evaluations to allow researchers and civil society to better understand how these systems work and the risks of bias and discrimination.

d) **Sharing information** on posts flagged by automated content moderation processes and systems; active bystanders that report content but are not themselves targeted by abuse; and users targeted by abuse.[209]

e) **Exercising 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.[210] For example, this could mean only sharing identity-related data at a population level (such as what percentage of reports were made by men vs women), and not disclosing details of a specific report that could identify people involved. When sharing involves personal data, services will need to consider data privacy risks and comply with the requirements of data protection law.[211]

f) **Ensuring that published information and findings are clear and accessible** to a range of audiences by producing versions of reports tailored to different audiences including children and young people. This should also include making sure information is well contextualised and explained so people are able to accurately interpret the findings.[212]

---

[208] Appelman, N., 2023. Disparate Content Moderation Mapping Social Justice Organisations Perspectives on Unequal Content Moderation Harms and the EU Platform Policy Debate. [accessed 4 November 2025].

[209] Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive. [accessed 29 October 2025].

[210] eSafety Commissioner, 2024. Technology, gendered violence and Safety by Design. [accessed 4 November 2025].

[211] For this good practice step and the steps above, more information on data privacy risks and complying with the requirements of data protection law can be found in the ICO UK GDPR guidance. See also the ICO's Data sharing: a code of practice.

[212] Service providers should have regard to the needs of their UK user base in considering what languages are needed to ensure published information and findings are clear and accessible.

4. Preventing harm

Overview

Context

4.1 Harm prevention means taking action to minimise the risk of harm before it occurs. To prevent harm, service providers should focus on deterring and interrupting perpetrator behaviour, including identifying potential routes for abuse and designing safety into features and functionalities. This draws on safety-by-design concepts that anticipate how features and functionalities can be co-opted to harm women and girls online and changing them to make it harder for perpetrators to misuse them. Online gender-based harms are often addressed retrospectively through interventions after the harm has occurred, which largely rely on women and girls to act, for example by reporting abuse to the service provider or the police. This requires time and effort from survivors and victims who are already navigating the psychological, emotional, and reputational effects of online gender-based harms.[213]

4.2 This chapter looks instead at the actionable ways providers can prevent and minimise online gender-based harms before they happen on their services, including by discouraging perpetrators from doing harm in the first place. We specifically look at three actions related to service design and harm prevention:

- **Action 4:** Conduct abusability evaluations and product testing.
- **Action 5:** Set safer defaults.
- **Action 6:** Reduce the circulation of content depicting, promoting or encouraging online gender-based harms.

4.3 For each action, we set out our expectation of what a baseline of safety looks like (**'foundational steps'**) for service providers to meet their duties to protect UK users. We have a range of enforcement powers to hold companies to account where they fail to comply with duties.

4.4 We also highlight additional **good practice steps** to illustrate how providers can build on the foundational steps to evaluate their systems, improve service design, and reduce harm.

4.5 These steps require a proactive approach which uses insights and evidence from risk assessments (see Chapter 3) to make changes to the features and functionalities of the service. As part of these risk assessment processes, providers may identify features or functionalities which were not made with a safety-by-design approach and therefore may need to be redesigned to be safer. This could include retiring certain features or introducing design changes to mitigate risks.

4.6 Many of these design changes facilitate safer behaviour or deter harmful behaviour through shaping 'choice

[213] Ofcom Stakeholder Workshop 1 on Women and Girls Online Safety, 16 September 2024.

4.7 architecture'. Choice architecture is the way service providers structure and present options to users—through design features, defaults, prompts, and interface layout—that shape how people make decisions online.[214]

4.8 While it is important to adopt this preventative approach when designing and integrating new technologies, for example GenAIchatbots[215] or virtual/augmented reality, it is also important to extend this approach to everyday digital technologies (social media, messaging, dating apps, search). [216] These sites are where the vast majority of online gender-based harms take place.[217]

Our target outcomes

4.9 Addressing the issue of online gender-based harms systematically means this Guidance is not focused on identifying individual instances of harm. Rather, online gender-based harms are structural issues which require active correction within various systems, from technology design to education. This preventative and systemic approach should reduce the burden on women and girls to keep themselves safe on services that have not been designed with their safety in mind. It can also limit the chilling effect for users witnessing gender-based harms, such as coordinated harassment, even if they are not the target.[218]

4.10 Safety-by-design can make features and functionalities harder to abuse, leading to a reduction in harms on the service.

4.11 Media literacy is also a vital part of preventing online gender-based harms.[219] By equipping people with the skills to critically engage online, we can prevent harm and foster a more resilient digital community in the UK. We include relevant examples of how service providers can promote media literacy as a preventative tool in this chapter.[220]

4.12 This approach includes engaging with men and boys to tackle misogynistic narratives and cultural norms that justify and glorify violence and abuse.[221] Providers can support this education through providing supportive or deterrence messaging, and ensuring they do not

---

[214] Ofcom, 2024. Understanding Online Choices, Preferences, and Welfare.

[215] The Act applies to certain types of GenAI content, chatbots and services. See Ofcom's open letter to online service providers which outlines how the Act will apply to Generative AI and chatbots.

[216] Mundane and everyday technologies are technologies which are used so commonly that they do not generate interest, excitement or attention unless they malfunction or are misused. For more information on mundane technologies, see: Dourish, P., Graham, C., Randall, D. and Rouncefield, M., 2010. Theme issue on social interaction and mundane technologies. *Personal and Ubiquitous Computing*, 14. [accessed 4 November 2025].

[217] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T. and Dell. N., 2018. A Stalker's Paradise": How intimate partner abusers exploit technology, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. [accessed 31 October 2025]; Refuge, 2022. Marked as Unsafe. [accessed 30 October 2025].

[218] The Global Partnership, 2022. Technology-Facilitated Gender-Based Violence: Preliminary Landscape Analysis. [accessed 29 October 2025].

[219] Ofcom Stakeholder Workshop 1 on Women and Girls Online Safety, 16 September 2024.

[220] See Ofcom consultation: Recommendations for how online platforms, broadcasters and streaming services should promote media literacy. Providers should refer to the final recommendations once the consultation has closed and these recommendations have been published.

[221] Burrell, S. 2018. The contradictory possibilities of engaging men and boys in the prevention of men's violence against women in the UK, *Journal of Gender-Based Violence*, 2 (3). [accessed 4 November 2025]; National Education Union, 2023. Working with boys and young men to prevent sexism and sexual harassment. [accessed 11 August 2025].

erroneously remove content which challenges the normalisation or promotion of misogynistic abuse and sexual violence (this can be referred to as 'counterspeech').

4.13 Preventative approaches are particularly valuable because they are often non-punitive, focusing on deterrence and behaviour change rather than punishment.[222]

4.14 It is also important to tailor these approaches to a variety of different kinds of users who may cause harm, from those that may do so unintentionally or unknowingly, to persistent and highly motivated perpetrators who will try to evade all safety measures. It is also important to tailor these approaches to different service types – for example, a search service might focus on reducing the risk of users encountering something they didn't want to see, and deterring users actively trying to use the service to find illegal or violative content. For a social media service or forum, the focus may also include deterring users from posting content they shouldn't, because it's illegal or violates terms of service.

Action 4: Conduct abusability evaluations and product testing

4.15 Abusability evaluations draw on the concept of 'usability' which is used to evaluate how easy it is for users to navigate a website or device to accomplish their goals. In contrast, abusability evaluations test how easy it is to abuse a tool or feature for harm, and therefore point to ways that abusability can be minimised in design.[223]

4.16 Perpetrators of online gender-based harms can be very innovative in co-opting technologies to facilitate pre-existing patterns of coercion and control. Many perpetrators – such as those spreading gendered and sexualised disinformation – use tactics to bypass detection and safety measures meant to prevent abuse, for example, through intentional misspellings of slurs and insults.[224]

4.17 This section outlines product testing methods that services can use that seek to anticipate these evolving forms of abuse (see Chapter 3). One way to do this is through 'red teaming' exercises, a practice that originated in cybersecurity but is now frequently used in other areas. Participants in the red team commonly take on the role of a bad actor to identify vulnerabilities in a system. They may also stress-test the robustness of safety features.[225]

4.18 Applying the concepts of abusability or red teaming during product testing can be useful across all forms of online gender-based harms. To maximise the effectiveness of red

---

[222] Ofcom Stakeholder Workshop 1 on Women and Girls Online Safety, 16 September 2024.

[223] Beers, A., Nguyên, S., Sioson, M., Mayanja, M., Ionescu, M., Spiro, E. S., and Starbird, K., 2021. The Firestarting Troll, and Designing for Abusability. [accessed 4 November 2025]; Ofcom Stakeholder Workshop 1 on Women and Girls Online Safety, 16 September 2024.

[224] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S. and Kaufmann, Z.), 2021. Malign Creativity: How Gender, Sex, and Lies are Weaponized Against Women Online. [accessed 29 October 2025].

[225] However, it is challenging to red team for some types of illegal content. It is a criminal offence under UK law to possess, show, distribute or make child sexual abuse material, meaning that directly red teaming for this content in the UK would risk someone becoming liable to prosecution. For more information on offences relating to CSAM, see the Illegal Content Judgements Guidance (section A5). However, the Government has announced upcoming legislation that will allow designated bodies to scrutinise AI models to check they cannot be exploited to generate CSAM, extreme pornography and non-consensual intimate images. Source: Department for Science, Innovation and Technology, Liz Kendall MP, and Jess Phillips, MP, 2025. New law to tackle AI child abuse images at source as reports more than double. [accessed 12 November 2025].

teaming, those conducting the tests should be familiar with the specific nuances and dynamics of online gender-based harms. Therefore, it is valuable to partner with subject-matter experts who have experience of supporting survivors and victims when designing and running such exercises.

Foundational steps: What are the expectations for service providers?

4.19 Our Codes and risk assessment guidance set out the following steps for service providers based on functionality, risk and size:[226]

a) **Product testing**: As a part of suitable and sufficient risk assessments (as outlined in Chapter 3), product testing is one of the types of evidence service providers could use as an input to improve the accuracy of their judgements on risk.[227]
b) **Significant change risk assessment**: Service providers must carry out a new risk assessment before making a significant change to their service.[228]
c) **Recommender system testing:** User-to-user service providers should, when carrying out existing on-platform testing of content recommender systems, collect additional safety metrics when making design adjustments, to evaluate whether the adjustment is likely to increase user exposure to illegal content.[229]

Good practice steps: How can service providers go further?

4.20 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps, both before and after product deployment, to gain further insights and data on the potential risks or weaknesses in their products. This will not only create safer environments for women, girls and other users at heightened risks of online gender-based harm, but in pre-empting and limiting misuse, providers can reduce the resource and reputational risks from abuse of their service by perpetrators.

4.21 The good practice steps providers could take include:

a) **Using red teaming** for abusability testing, in which a team takes on the role of a malicious actor and tries to find vulnerabilities in a system (see **Case Study 8**).[230] The team could be internal, or include external subject-matter experts. This could be repeated periodically and iteratively, even if there are no major developments, as perpetrators will adapt quickly to evade safety measures.
b) **Using personas** to explore how different users may experience a feature, map user journeys, and increase understanding of intersectional perspectives.[231]

---

[226] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services.
[227] Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance on "core evidence" and "enhanced evidence" for risk assessments.
[228] Illegal Content Risk Assessment Guidance and Children's Risk Assessment Guidance.
[229] Illegal Content Codes of Practice (ICU E1). This measure only applies to user-to-user services that identify as medium or high risk for at least two specified harms.
[230] Ofcom, 2024. Red Teaming for GenAI Harms - Revealing the Risks and Rewards for Online Safety.
[231] World Wide Web Foundation, 2021. Tech Policy Design Lab: Online Gender-Based Violence and Abuse. [accessed 30 October 2025].

c) **Working with experts** with direct or relevant experience engaging with and understanding perpetrator behaviours. As perpetrator tactics evolve quickly, engaging with experts on a regular basis and staying up to date with research on perpetrator behaviour is invaluable (see **Case study 3**).
d) **Quality assurance**: Evaluating methods and outcomes of product testing to understand and ensure the efficacy of these methods.
e) **Media literacy:** Adhering to the principles of monitoring and evaluating features in the Best Practice Design Principles for Media Literacy.

4.22 **Case study 7** sets out examples of how a social media service provider can conduct abusability testing to prevent pile-ons and coordinated harassment on its service.

Case study 7 (good practice): Abusability product testing

Scenario

A **social media provider** plans to introduce a feature which allows creating and sharing lists of other users on the service. The provider knows that similar features on other services have been misused to share lists of public figures with specific characteristics as targets for **pile-ons and coordinated harassment** and want to avoid exposing its users to these harms.

Steps to take

The provider decides to **undertake abusability testing** for its new feature:

- The provider sets up an **abusability evaluation** for the prototype feature and tests whether the feature could enable a pile-on. It involves in-house experts, as well as experts from organisations that represent people with lived experience of coordinated harassment (see **Case study 1**).
- This helps the provider identify 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. When the feature is launched, users can mitigate the risk of harm from such features with user controls designed with the harm of pile-on and coordinated harassment in mind.

Considerations

- Providers should carefully weigh risks when making design choices. In some cases, removing such a feature could remove the risk of misuse without substantially inconveniencing legitimate users. In other 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.[232]

[232] Freed, D., Palmer, J., Minchala, D., Levy, K., Ristenpart, T. and Dell, N., 2018. "A Stalker's Paradise": How Intimate Partner Abusers Exploit Technology, *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* [accessed 28 October 2025]. Parkin, S., Patel, T., Lopez-Neira, I., and Tanczer, L., 2020.; Slupska, J. and Tanczer, L., 2021. Threat Modelling Intimate Partner Violence: Tech Abuse as a Cybersecurity Challenge in the Internet of Things in *The Emerald International Handbook of Technology-Facilitated Violence and Abuse.* [accessed 22 October 2025].

4.23 **Case Study 8** sets out examples of how a general search provider can use red teaming to prevent deepfake intimate image abuse.

Case study 8 (good practice): Red teaming for harmful user queries

Scenario

A large general search provider integrates a **GenAI summary** feature on its service to produce a text-based summary of selected search results for the user.

Although the provider has already introduced moderation steps to ensure nudification sites are not surfaced in organic search results, (e.g. "blue link" results) (see **Case Study 13** on gender-sensitive search services), specialist organisations are increasingly reporting that GenAI tools can produce instructions on how to generate **deepfake intimate image abuse**.

The provider is concerned that its GenAI summary feature could have similar vulnerabilities, which may undermine the effectiveness of the moderation steps (i.e., allowing users to effectively circumvent those measures to access content that would otherwise be restricted under the provider's internal content policies for the service).

Steps to take

As part of its **red teaming exercise** to help identify vulnerabilities in the GenAI summary feature, the provider decides to test whether the feature may produce or enable a user to access instructions on how to create deepfake intimate image abuse content.

- When setting up this exercise, the provider considers the best practices set out in Ofcom's discussion paper on red teaming, which include making sure that red teaming participants have relevant subject-matter expertise on online gender-based harms.
- Participants in the red team adopt various techniques to assess whether the GenAI summary feature can return harmful content for women and girls. For example, misspelling words in order get step-by-step instructions on how to generate deepfake intimate image abuse.
- Once it has analysed the results of the exercise, the provider decides on the most appropriate steps to take, which includes adopting new safety measures. The provider introduces a keyword-based safety filter to prevent the GenAI summary feature from returning responses that may facilitate gender-based harm and updates the list of blocked websites for the model.
- The provider also ensures it is documenting and sharing the results of red teaming tests with relevant internal teams.

Considerations

- It is a criminal offence under UK law to possess, show, distribute or make CSAM or attempt to do so.[233]
- As per every other form of testing and evaluation, red teaming will not necessarily discover every vulnerability in the tool. While red teaming is meant to emulate how users would interact with a model in real life, there are infinite ways people can use these tools. Bad actors will also attempt to override safeguards by changing circumvention methods. This means that red teamers will not always be able to discover every vulnerability or mirror the behaviours of bad actors in a timely way.

[233] See footnote 229 for more information on red teaming for CSAM.

Action 5: Set safer default settings

4.24 A common outcome of abusability evaluations will be to set safer defaults. Default settings can be a powerful tool to create a safer experience online, particularly when they are complemented with the harm prevention methods set out in Action 6. Our research shows that defaults (such as a pre-selected choice) strongly influence user choice even when changing the setting takes one click.[234]

4.25 Safer defaults can also embed better consent practice in service design by allowing for more points where users can determine whether they want to participate in an interaction or allow their data to be shared.

4.26 Taking steps to make the service less susceptible to abuse by default makes it easier for all users to keep themselves safe online. In doing so, this should reduce the burden of 'safety work' (such as having to report many individual pieces of content), experienced by survivors and victims of gender-based harms.

4.27 These safety measures often cut across different types of harm, although they are particularly important for harms like stalking and coercive control.

Foundational steps: What are the expectations for service providers?

4.28 Our Codes set out the following steps for service providers based on functionality, risk and size:[235]

a) **Safe settings**: User-to-user services' default settings should protect child users.[236]

i) Automated location information of child users' accounts should not be visible to any other users by default. Any location sharing functionality should be 'opt in'.
ii) Child users should not be visible in connection lists of other users. The connection lists of child users should also not be visible to other users.
iii) Child users are not presented with prompts to expand their network of friends, or included in network expansion prompts presented to other users.
iv) Non-connected accounts do not have the ability to send direct messages to children using a service.
v) For services with no formal connection features, providers should introduce mechanisms to ensure children using a service can actively confirm whether to receive a direct message sent from another user account before it is visible to them.

b) **Group chats**: User-to-user service providers should provide children the option to accept or decline an invitation to a group chat. [237]

---

[234] Ofcom, 2024. Behavioural insights to empower social media users.
[235] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[236] Illegal Content (ICU F1).
[237] Protection of Children (PCU J3).

c) **Supportive information**: User-to-user service providers should provide children supportive information in a timely and accessible manner.[238] This is to help child users make informed choices about risk by giving them information, access to safeguarding processes, and support on a service, when they are:

   i) seeking to disable one of the safer defaults recommended previously which are set to reduce risk;
   ii) responding to a request from another user to establish a formal connection; and
   iii) receiving a direct message from another user for the first time.

d) **Safe search**: Providers of large general search services likely to be accessed by children should filter out identified pornography, eating disorder, suicide and self-harm content (or a relevant proxy of such content)[239] for any users they have determined are child users.[240]

Good practice steps: How can service providers go further?

4.29 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps to ensure that safer defaults benefit all users, not only children. When the highest safety option is pre-selected by default, providers can demonstrate their commitment to safety by ensuring women and girls have a safe and accessible online experience from their first moment on the service.

4.30 The good practice steps providers could take include:

a) **Clear communication:** Clearly explaining default settings, bundles, and account access options to all users, for example through visual elements, audio-visual elements, or interactive elements which are comprehensive and easy for all users to understand. This is particularly important when default settings change, or when a user turns default settings off.
b) **Interaction defaults:** Setting strong and customisable defaults for all users (both adults and children) around interaction, such as who can contact a user or asking a user's permission before being added to a group chat. Adults should have the ability to change these settings.[241]
c) **Privacy defaults:** Setting strong and customisable defaults around privacy, such as what information users can see about another user, including their location (see **Case study 9**) , their content and who can and cannot redistribute their content or username/identity in real time.[242] Other settings could include providing users with

[238] Illegal Content (ICU F2).
[239] By 'proxy' we mean content that a provider determines to be in breach of its terms of service, where: a) the provider had reason to suspect that the content may be relevant primary priority content and b) the provider is satisfied that its terms of service prohibit the type of relevant primary priority content which it had reason to suspect existed.
[240] Protection of Children (PCS C2). For this measure, we set out that where providers choose to filter out such content via a functionality that applies by default for users, (such as a 'safe search default'), providers should take steps to ensure that this functionality or safe search setting cannot be switched off by users determined to be children.
[241] Under upcoming User Empowerment duties, Category 1 services must give adults the options to change the default setting of a control feature, whether it is on or off by default (see section 15(5) and (6)) of the Act. However, service providers may still choose to make some safety settings (such as privacy defaults around who can reshare a child user's content) mandatory for child users.
[242] For more information on default privacy settings, see the ICO's Age Appropriate Design Code of Practice, Chapter 7: Default Privacy Settings.

tools to express their gender identity and giving them control over the provider's collection and inference of user information related to their sexual orientation and gender identity.[243] Adults should have the ability to change these settings.[244]

d) **Notification settings:** When designing notification settings about user interactions (e.g. tagging, screenshots and screen recordings, location sharing), consider the potential risks and benefits for people experiencing gender-based harms, in particular stalking and coercive control (these could be surfaced through abusability testing). For example, survivors and victims of stalking and coercive control might want to be informed when they are tagged in another user's content, particularly if the latter has less private visibility settings, and they prefer for other users not to be informed about their online activity or location.

e) **Bundles:** Combining relevant interaction and privacy settings into 'bundles'. Users can be overwhelmed by too many options for settings, including when creating a profile, meaning they disengage from making decisions. Grouping relevant choices together into a 'bundle' can reduce the time and effort required from users and customisation options can help users make informed choices about their online settings.[245] On the other hand, offering customisation options could allow users to make more granular choices outside of the bundles. Poorly designed bundles could undermine user agency, especially for vulnerable users, or increase the risk that users are not informed about data processing, so bundles should always be clearly explained and assessed for compliance with data protection and impact on privacy.[246]

f) **Strengthening account security** with Two Factor Authentication ('2FA') or multi-factor authentication, while ensuring accounts can be recovered after being hacked and/or locked out. ICO guidance highlights that this will be more important where personal data that can be accessed is of a sensitive nature or could cause significant harm if it were compromised.[247] It is important for this to be paired with providing users with advice about protecting their data,[248] and for providers to be aware of how security measures can be bypassed by motivated perpetrators.[249]

g) **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 devices (via IP/MAC addresses) are connected to an account, and making it easy to collect evidence of illicit account access.[250] This minimises opportunities for non-consensual monitoring and surveillance in the context of stalking and coercive control,

---

[243] GLAAD, 2024. GLAAD Social Media Safety Index Platform Scorecard. [accessed 29 October 2025]; Department for Science, Innovation & Technology and PUBLIC, 2025. Platform Design and the Risk of Online Violence Against Women and Girls. [accessed 29 October 2025].

[244] See footnote 241.

[245] Behavioural Insights Team, 2021. Active Online Choices: Designing to Empower Users. [accessed 29 October 2025].

[246] Data protection law requires that providers are clear and open with users about how their personal data is used. See ICO guidance on the right to be informed and transparency. See also the ICO's Guidance on default settings and valid consent under UK GDPR, and the ICO and CMA's joint paper on harmful design.

[247] ICO guide to data security.

[248] ICO, Passwords in online services. This also highlights the importance of ensuring that any processing of biometric data for the purpose of uniquely identifying an individual is done in accordance with the requirements for processing special category personal data in line with data protection legislation.

[249] Leitão, R., 2019. Anticipating Smart Home Security and Privacy Threats with Survivors of Intimate Partner Abuse. [accessed 26 August 2025].

[250] Service providers should have regard to the needs of their UK user base in considering what languages are needed to ensure the information about account access is accessible.

and supports survivors and victims in putting together a case if nonconsensual monitoring occurs.[251]

h) **Reminders:** Identifying optimal frequency and timing and giving users regular reminders for reviewing or updating privacy and security settings, with easy-to-understand language.[252] Research finds that timing of prompts matters for user engagement, and overly frequent reminders can be annoying, leading to disengagement.[253]

i) **Engaging with subject-matter experts**, particularly those with experience of supporting survivors and victims, when designing privacy and security settings (**Case Study 3**).[254]

4.31 **Case study 9** sets out examples of how a social media provider can set safer defaults related to location data for all users to prevent stalking.

Case Study 9 (good practice): Location information on social media messaging

Scenario

A **social media provider** receives reports that users have been located by perpetrators of **stalking** through location data attached to photos and videos. The provider realises most users are not aware that this metadata is exposed through uploading content.

Steps to take

The provider decides to switch off geolocation options by default, to limit opportunities for accidental location sharing.[255] The provider:

- Alerts users when location tracking is on and links to information on what data this shares with other users.
- Retains the option for location metadata to be re-enabled, but provides clear, comprehensible and easy to understand information alongside this setting, making users aware that their location may be seen by other users.

Considerations

- When using default settings, providers should test their effect to identify any new risks that they might inadvertently introduce (see **Case study 7** on abusability testing) and be mindful that accepting a default is not the same as making an active choice. They should also be aware that defaults might need to change over time, to accommodate new terms, policies or technology and relevant data protection laws.

---

[251] When introducing this good practice step, services must ensure that the underpinning personal data processing is adequate, relevant, and limited to what is necessary, in compliance with data protection law. See the ICO Guidance on data minimisation.

[252] Service providers should have regard to the needs of their UK user base in considering what languages are needed to ensure reminders are accessible.

[253] Wohllebe, A., Hübner, D., Radtke, U. and Podruzsik, S. 2021. Mobile apps in retail: Effect of push notification frequency on app user behavior. *Innovative Marketing* 17 (2). [accessed 30 January 2025].

[254] Providers should also consult resources from the ICO which will support organisations to consider data privacy and security.

[255] Ensuring geolocation options are off by default aligns with standard 10 'Geolocation' of the ICO's Age Appropriate Design Code, which helps providers of online services where they are likely to be accessed by children take necessary steps to protect their personal data.

Action 6: Reduce the circulation of content depicting, promoting or encouraging online gender-based harms

4.32 Providers can also prevent harm by using techniques that reduce the risk of individuals encountering illegal content (such as intimate image abuse and cyberflashing),[256] protect children from encountering content harmful to them (such as abuse and hate).[257] Such techniques can also help providers to enforce the policies they choose to set out in their terms of service (see Chapter 3).

4.33 Some providers may choose to take steps to protect all users, not just children,[258] from encountering harmful content, for example because it violates their terms of service. Ultimately, it is up to providers to decide what content they allow on their services as long as this complies with the duties set out in the Act. This includes choosing if they want to allow adults to post and view content that does not constitute a criminal offence, but may be abusive, hateful or violent. If providers choose to allow or promote such content, they should be transparent with users about the choices they make and content that users may be exposed to.[259]

Foundational steps: What are the expectations for service providers?

4.34 Our Codes set out the following steps for service providers based on functionality, risk and size:[260]

a) **Signposting users to support:** User-to-user service providers should signpost children to support when they post harmful content, including bullying, suicide, self-harm or eating disorder content, or search for suicide, self-harm or eating disorder content on a user-to-user service.[261] Providers of large general search services should provide users with crisis prevention information in response to search requests regarding suicide, self-harm and eating disorders.[262]

b) **CSAM warnings for search:** Search service providers should have systems and processes to detect and provide content warnings and support resources for users making search requests where the wording clearly suggests that the user may be seeking to encounter

[256] The safety duties for user-to-user service providers about illegal content are set out in section 10 of the Act. Duties for search service providers about illegal content are set out in section 27 of the Act.
[257] The safety duties for user-to-user service providers about protecting children are set out at section 12 of the Act. Duties for search service providers are set out at section 29 of the Act.
[258] Duties relating specifically to the protection of children are set out in sections 11-13 and 20-21 of the Act for regulated user-to-user services and sections 28-30 and 31-32 of the Act for regulated search services.
[259] This includes providing information in their terms of service (for user-to-user services) or publicly available statements (for search services) about how children will be protected from content harmful to them, and for Category 1 and 2A services summarising the findings of their most recent children's risk assessment. See sections section 12(9), 12(14), 29(7) and 29(9) of the Act.
[260] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[261] Protection of Children (PCU F4, PCU F5).
[262] Illegal Harms (ICS F3) and Protection of Children (PCS F3).

CSAM, or which use terms or combinations or letters and symbols that explicitly relate to CSAM.[263]

c) **Highly effective age assurance**: User-to-user service providers should use highly effective age assurance to prevent and/or protect children from encountering content harmful to children which they do not prohibit on the service, or which they prohibit but are unable to remove, and/or to apply the relevant recommender system measures (see below).[264]

d) **Hash matching:** User-to-user service providers should use the automated technique known as 'hash matching' to detect and remove image-based CSAM.[265]

e) **Content moderation**: User-to-user service providers should use content moderation to review and assess suspected content harmful to children and suspected illegal content, and act swiftly when they identify such content where it is currently technically feasible.[266]

f) **Search moderation:** Search service providers should have systems and processes designed to review, assess and where relevant take appropriate moderation action in relation to search content which the provider has reason to suspect may be illegal content or content harmful to children.[267] In addition, service providers should ensure that users do not encounter search content present at or sourced from URLs or domains[268] previously identified as hosting CSAM and that are included on a list of URLs sourced from a person with expertise in CSAM identification.[269]

g) **Recommender systems**: User-to-user service providers should ensure that any content recommender system that children can access is designed and operated so that content indicated potentially to be primary priority content – pornography, eating disorder content, self-harm and suicide content – is excluded from children's recommender feeds.[270] In addition, content that is indicated potentially to be priority content, including abuse on the basis of sex or gender reassignment, as well as content promoting gendered violence and depression and stigmatising body types, or an identified kind of non-designated content, should be excluded from or given a low degree of prominence in children's recommender feeds (see **Case study 10**).[271]

---

[263] Illegal Harms (ICS F2).

[264] Protection of Children (PCU B1, PCU B2, PCU B3, PCU B4, PCU B5, PCU B6, and PCU B7). In our Protection of Children Code of Practice for user-to-user services, we outline which user-to-user services should use highly effective age assurance to (i) prevent children from accessing the entire service where their principal purpose is to host or disseminate primary priority content or priority content; (ii) to ensure children are prevented from encountering primary priority content identified on the service; (iii) to ensure children are protected from encountering priority content identified on the service; and/or (iv) to apply relevant recommender system measures to children. We set out what is 'highly effective age assurance' in our Part 3 Guidance on highly effective age assurance, published in April 2025.

[265] Illegal Harms (ICU C9).

[266] Protection of Children (PCU C1 and C2). For a detailed discussion of content harmful to children, see the Introduction chapter and our Children's Register of Risks. [accessed 12 September 2025].

[267] Illegal Harms (ICS C1), Protection of Children (PCS C1).

[268] This includes search content present at or sourced from domains which are entirely or predominantly dedicated to CSAM.

[269] Illegal Harms (ICS C7).

[270] Protection of Children (PCU E1).

[271] Protection of Children (PCU E2).

4.35 **Case study 10** sets out examples of how a video sharing service provider can take steps to prevent children from encountering misogynistic abuse and sexual violence.

Case study 10 (foundational): Evaluating recommender systems to tackle misogynistic abuse and sexual violence

Scenario

A **video sharing service provider** becomes aware through a youth organisation that many young boys are seeing content depicting misogynistic abuse and sexual violence. After an initial testing session where it mirrors the experience of a young boy on its service, it discovers that videos about building confidence often lead to videos which promote misogynistic abuse and sexual violence being promoted to them through their recommender algorithm.[272]

Steps to take

The video sharing service provider conducts and records an evaluation of how its recommender systems surface content promoting misogynistic abuse and sexual violence , to children.

- After this evaluation, the provider takes action to retrain its recommender algorithm. It improves techniques to identify and reduce the prominence of misogynistic abuse and sexual violence.
- While implementing the changes, the provider keeps a record of the evaluation in order to document its decisions for future risk assessments (see Action 2) and to share learnings internally.

Considerations

- As well as misogynistic abuse and sexual violence, providers should consider the other types of content indicated potentially to be priority content that should be excluded from or given a low degree of prominence in children's recommender feeds. They should also exclude content indicated potentially to be primary priority content.
- The provider could go further to consider introducing nudges to discourage users from uploading content promoting misogynistic abuse and sexual violence.
- During this development process, the provider considers the risk of user friction and fatigue in deciding what interventions are appropriate. In addition to evaluating and improving recommender systems, platforms can engage with subject-matter experts (see **Case Study 3**) who support and work with parents, young people and carers. Media literacy interventions in schools, communities and at home can promote an understanding of how services and influencers' financial incentives shape the content we see.[273]
- The provider should have regard to the ICO's Guidance on AI and Fairness which sets out a clear methodology for auditing AI applications.[274]

---

[272] SAFER SCROLLING How algorithms popularise and gamify online hate and misogyny for young people [accessed 20 August 2025].
[273] Ofcom / Young People's Action Group Roundtable, 7 July 2025; For more information on media literacy, please see Ofcom's Best-Practice Principles for On-Platform Interventions to Promote Media Literacy. [274] ICO, Guidance on AI and Fairness.
[274] ICO, Guidance on AI and Fairness.

Good practice steps: How can service providers go further?

4.36 In addition to the foundational steps, there are other techniques that providers can use to reduce the impact of harmful content on women and girls. These good practice steps give service providers the opportunity to innovate and gather data on new approaches – some of which we are actively exploring for future Codes of Practice. This could not only help to strengthen providers' responses to online gender-based harms, but also help them set themselves apart as industry leaders in this highly complex and vital space.

4.37 In this section, we explore three methods: Persuasion (introducing deliberate friction), Removal (preventing uploads or take downs) and Reduction (limiting circulation or reducing visibility). Many of these methods will rely on automated detection technologies. As is explored in the following sections in detail, our good practice suggests:

a) Using persuasion to address all four areas of online gender-based harms.
b) Using removal to address illegal gender-based harms (stalking, coercive control and image-based sexual abuse).
c) Using reduction to address gender-based harms that include content harmful to children (misogynistic abuse and sexual violence and pile-ons and coordinated harassment). We also look at how search services can reduce exposure to links that lead to or enable illegal harms.

4.38 Given these techniques impact how people can participate in online dialogues and view and discover different types of content, providers should carefully consider which good practice is most appropriate for their service type, and the kinds of harms they are targeting. This includes considering their duties in relation to freedom of expression and privacy (see Action 1).

4.39 Providers should also note that decisions they make about the circulation of online content will affect different groups in different ways. For example, reducing the visibility of content criticising women in public life will impact how far some users can spread their message. However, choosing not to restrict the circulation of such content where it amounts to a pile-on means that users (in this case, women in public life) may end up limiting their speech or withdrawing from public life.

4.40 The methods described for persuasion, removal and reduction often rely on content moderation systems to scan, identify, and filter content depicting online gender-based harms. Some providers may use automated detection, including the use of proactive technology, which Ofcom is only able to recommend in our Codes for content communicated publicly.[275] It is important to continuously improve automated systems to identify content that could be harmful and avoid recommending content, forums or groups that are likely to encourage misogynistic abuse and sexual violence. This could involve:

a) Evaluating automated systems to ensure they are accurate, effective, contextually nuanced and minimise bias in terms of race, gender, and other characteristics.[276] For example, automated systems should seek to avoid erroneously removing content by

[275] Ofcom is only able to recommend the use of proactive technology for the Illegal Harms, Protection of Children and Fraudulent Advertising Codes (see paragraph 13(3) of Schedule 4).
[276] Appelman, N., 2023. Disparate Content Moderation Mapping Social Justice Organisations Perspectives on Unequal Content Moderation Harms and the EU Platform Policy Debate. [accessed 22 October 2025].

survivors and victims that calls attention to racism and misogynistic abuse (see **Case study 14**).[277]

b) Training detection algorithms to account for culturally sensitive context and develop multi-modal systems capable of greater nuance.[278] In particular, emerging research suggests large language models can be effective at incorporating context in hate speech detection, but raise new challenges related to biases and consistency, such as misclassifying content from marginalised communities as false positives.[279]

c) Ensuring that moderation is effective for different kinds of formats like text, image, and voice. In-game voice chat moderation can also reduce exposure. For example, machine-learning driven voice moderation can detect harmful audio content.[280]

d) Adding safeguards for freedom of expression, including routes to contest wrongly moderated content. Hybrid content moderation systems which combine automation with human moderators for challenging, context dependent cases can help improve outcomes (see Action 9).[281]

Persuasion

4.41 Persuasion refers to introducing friction into the user journey to encourage a user not to post or upload online gender-based harms. This aims to give people time to think about what they are posting. These methods are beneficial to educate users about respectful behaviour.

4.42 As persuasion does not block content from being uploaded, it is a useful technique to address misogynistic abuse and sexual violence which includes both illegal content and content harmful to children. Persuasion can also deter perpetrators of intimate image abuse. Persuasion is less likely to be effective against highly motivated perpetrators of illegal harms, for example in cases of stalking and coercive control.

---

[277] Kwarteng, J., 2022. Misogynoir: Challenges in Detecting Intersectional Hate, *Social Network Analysis and Mining,* 12 (1). [accessed 31 October 2025].

[278] Peterson-Salahuddin, C., 2024. Repairing the harm: Toward an algorithmic reparations approach to hate speech content moderation, *Big Data & Society*, 11 (2). [accessed 23 October 2025]; Chan, A. J., Redondo García, J. L., O'Donnell, C., Silvestri, F. and Palla, K., 2024. Enhancing Content Moderation with Culturally-Aware Models. [accessed 23 October 2025]; Arya, P., Pandey, A. K., Patro, S. G. K., Tiwari, K., Panigrahi, N., Naveed, Q. N. and Khan, W. A., 2024. MSCMGTB: A Novel Approach for Multimodal Social Media Content Moderation Using Hybrid Graph Theory and Bio-Inspired Optimization, *IEEE Access* (12). [accessed 23 October 2025]; Wang, W., Huang, J., Huang, J., Chen, C., Gu, J., He, P. and Lyu, M. R., 2023. An Image is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software. [accessed 23 October 2025].

[279] Albladi et al. 2025. Hate Speech Detection Using Large Language Models: A Comprehensive Review. [accessed 29 August 2025].

[280] Pappas, M., 2023. Social Safety in Games: Moderating Voice Chat in the Metaverse, *ACM Games: Research and Practice*, 1 (3). [accessed 31 October 2025].

[281] Albladi et al. 2025. Hate Speech Detection Using Large Language Models: A Comprehensive Review. [accessed 29 August 2025].

4.43 Persuasion could include:

a) **Introducing deliberate friction** through prompts, alerts, and nudges[282] which can interrupt actions and encourage reflection. This can be useful in the context of:

i) Asking users to reconsider when posting misogynistic abuse and sexual violence (see **Case study 11**).[283] Prompts can also be used to deter perpetrators from sharing non-consensual intimate content. For example, providers can use nudity detection to flag when a user is about to share intimate images and check if they are sure they want to share, reminding them it is illegal to share an intimate image of someone without their consent.

ii) Deterring users searching how to perpetrate stalking, coercive control, or intimate image abuse, where providers can identify search queries directly related to these harms. This can be done by prioritising search results that highlight where such content or activity is illegal and refer users to support services instead. This deterrence method can also be used where a search service has integrated a large language model (LLM) to generate text-based summaries of search results and users are inputting search queries related to illegal harms.[284]

b) **Allowing people to use identity verification:** Giving users the choice to verify their identity, or only engage with other verified users, may reduce the online disinhibition effect.[285] This effect sees individuals engaging in harmful behaviour like misogynistic abuse, pile-ons and intimate image abuse that they would typically refrain from engaging in offline because of anonymity or pseudonymity.[286] While identity verification may be useful in some circumstances (see **Case study 12**), effective and appropriate implementation is key given the privacy implications, especially as people may seek to be anonymous/pseudonymous online for a range of reasons, including to avoid stalking and to access supportive resources.[287] Service providers should consider the data protection implications related to identity verification.[288]

---

[282] Nudges are design measures within an online environment that promote some behaviours and/or discourage others.

[283] Cox, A. L., Gould, S. J., Cecchinato, M. E., Iacovides, I. and Renfree, I., 2016. Design frictions for mindful interactions: The case for microboundaries, *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. [accessed 31 October 2025].

[284] For more information on how the Act applies to GenAI tools, see Ofcom's open letter.

[285] More evidence is required to establish the effectiveness of identity verification in reducing the online disinhibition effect. However, evidence suggests that the use of anonymous profiles enables users to do or say things online that they may not do in person, encouraging them to engage in hateful or abusive behaviour online. See further details on the risks of anonymity and pseudonymity in Ofcom Illegal Harms Register of Risks and Children's Register of Risks.

[286] Cheung, C.M., Wong, R.Y.M. and Chan, T.K., 2021 Online disinhibition: conceptualization, measurement, and implications for online deviant behavior, *Industrial Management & Data Systems*, 121 (1). [accessed 3 November 2025]; Ofcom Stakeholder Workshop 2 on Women and Girls Online Safety, 19 November 2024.

[287] We will be considering the details of identity verification further in a future consultation so Category 1 providers who implement this may wish to review this in due course. For further information, see Ofcom's approach to implementing the Online Safety Act.

[288] Service providers can use the ICO's UK GDPR guidance and resources to ensure their approach to user verification is necessary, proportionate and compliant with data protection law. Service providers must follow the key principles of data protection law such as purpose limitation, data minimisation, and storage limitation. Providers must also identify a lawful basis to process personal data – see the ICO guidance on lawful bases for more information. If relying on consent as a lawful basis for identity verification, providers must ensure it meets the UK GDPR standards for validity.

4.44 **Case study 11** sets out examples of how a dating app can prevent misogynistic abuse by introducing prompts.

Case study 11 (good practice): Misogynistic abuse on a dating app

Scenario

A **dating app** receives increasing reports of users receiving abusive content, including unwanted sexual advances. In some cases, this unsolicited contact escalates into a persistent and repeated pattern of behaviour. The service provider wants to better deter users from engaging in harmful behaviour and support users when the behaviour does occur.

Steps to take

The service provider decides to introduce prompts that interrupt harmful actions by adding friction and making potential harms more salient at the point of action.[289]

- When harmful or abusive language is detected in a message, a warning appears before the content is sent, giving the sender an opportunity to edit. The added friction slows impulsive behaviour by adding an extra step before a harmful action is taken, such as sending an unwanted sexual advance. Drawing attention to risks in the moment helps users recognise potential harm and pause before acting.
- Where patterns such as repeated messaging without response are identified, the service also issues cautionary prompts.
- These short, in-the-moment interventions reduce impulsive harmful behaviour by encouraging users to reconsider their actions. In turn, this preventative step reduces the likelihood of this behaviour on the app.

Considerations

- Prompts are most effective at curbing impulsive or "in the moment" behaviour, where users act without thinking. They are less likely to influence determined perpetrators who knowingly intend to cause harm.
- Prompts that are complex, highly technical, overly punitive, or shaming may alienate users,[290] and for children, risk creating fear or discouraging open communication with parents or guardians.
- Providers should also consider the risks of prompt fatigue – a phenomenon where too many pop ups or notifications may lead users to ignoring them – when designing such interventions. Over time, users may also become desensitised and click through prompts automatically. Rotating message style, wording, or design can help sustain their impact.
- Automated systems may miss more subtle or context-specific forms of abuse (e.g., patterns of stalking and coercive control, use of coded language). Prompts should therefore complement, not replace, other safeguards.
- Prompts are one layer of protection. They should be embedded within a broader safety ecosystem that includes effective moderation, reporting mechanisms, and enforcement

[289] Anderson, B.B., Kirwan, C.B., Eargle, D., Jensen, S.R. and Vance, A., 2015. Neural correlates of gender differences and colour in distinguishing security warnings and legitimate websites: a neurosecurity study. Journal of Cybersecurity, 1(1), pp.109-120. [Accessed 20 November 2025]

[290] Zaaba, Z.F., Lim Xin Yi, C., Amran, A. and Omar, M.A., 2021. Harnessing the challenges and solutions to improve security warnings: A review. *Sensors*, *21*(21), p.7313). [Accessed 20 November 2025]. Framing should be age-appropriate, supportive, and carefully tested.

actions. For example, the service provider could also consider introducing prompts on the opposite side of the user interaction, directing users towards reporting systems if abusive content is detected in a message they receive.

- Providers can also consult Ofcom's Best-Practice Principles for On-Platform Interventions to Promote Media Literacy when designing prompts. Interventions could aid understanding and increase literacy surrounding risks and behaviours that can prevent harm as well as around how to correct mistakes.

Removal

4.45 Removal refers to using technical tools to block uploads of content, or to remove content after it has been uploaded.[291]

4.46 The good practice below only covers the use of removal methods for illegal online gender-based harms (stalking, coercive control and image-based sexual abuse). Providers may choose to use removal more widely where content violates terms of service.[292]

4.47 This could include:

a) **Using hash matching** to prevent uploads of known intimate image abuse.[293]
b) **Introducing time-out features** to users who repeatedly attempt to use service features or functionality to perpetrate stalking and coercive control or intimate image abuse. This means the affected user would not be able to send a message or use other service features for a set amount of time while in time-out. If a user continues to misuse the service, repeated strikes could lead to an account ban (see Action 9).
c) **Requiring evidence of consent** from those depicted in intimate content prior to uploading where adult content is allowed on a service to prevent intimate image abuse, including deepfakes (see **Case study 12**).
d) **Introducing prompt and output filters** for GenAI models to stop them from generating intimate image abuse content or promoting stalking and coercive control.

4.48 **Case study 12** sets out examples of how an adult service provider can take steps to prevent image-based sexual abuse.

[291] In this section, we focus on methods for doing this automatically rather than in response to user reports. Content moderation following user reports is covered in Chapter 5.
[292] In this chapter, we do not specify what service providers' terms of service should regulate, but rather review how they can enforce the policies they set out in their terms of service (see Governance and Accountability in Chapter 3).
[293] At the time of publication, Ofcom's consultation on Additional Safety Measures has closed, and we are carefully considering responses. It includes the proposal to use perceptual hash matching to detect image-based intimate image abuse content so it can be removed (for some user-to-user services) or so it can be moderated (for providers of large general search services). Consultation: Online Safety - Additional Safety Measures - Ofcom.

Case study 12 (good practice): Preventing image-based sexual abuse on adult services

Scenario

A news investigation reveals that a **pornography service** with user-generated content has been hosting **intimate image abuse content**. The service provider realises it has not adequately evaluated the risk of this harm on its service. Providers of pornography services face higher risks of hosting intimate image abuse content because their services allow sexual content.[294]

Steps to take

The provider introduces a variety of persuasive user measures and additional removal features to mitigate this risk. The provider:

- Introduces user verification to the upload process for content uploaded by the service's creators, asking all individuals who are depicted in content uploaded to the service to provide a full legal name, date of birth, a piece of matching government-issued photo ID, and a go through a live face scan check. Introducing this process could be coupled with removing historic videos depicting unverified individuals.
- Introduces consent verification, where all individuals depicted in uploaded content must certify that they have consented to appear.
- Adds deterrence messaging to the upload process, i.e. warning messages about the illegality and consequences of intimate image abuse.
- Introduces better removal for survivors and victims, the provider also adds hash matching,[295] by joining a cross-industry initiative, such as StopNCII.org, which allows survivors and victims to generate hashes from their intimate images. These hashes are shared across participating service providers to detect and prevent the non-consensual circulation of these images.

Considerations

- Providers can layer different techniques to prevent intimate image abuse content. Service providers may also refer to the Image-Based Sexual Abuse Principles.
- For pornography services which allow studios to upload content at scale, providers can have contracts in place that require the studio to hold the consent and ID documents of all individuals in the content they upload. Pornography services can also use regular audits of the studios, to check this documentation, and have contract termination clauses related to breaches of this requirement.
- Identification and consent verification measures must be compliant with data protection law. Service providers that choose to follow this step should ensure that they take a data protection by design and default approach to their processing ensuring that necessary safeguards are integrated to protect the rights and freedoms of users. [296]

---

[294] Illegal Harms Register of Risks.

[295] 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.

[296] Service providers should consider a range of options for their service before determining whether the proposed approaches set out in the case study are an appropriate and proportionate means of taking the good practice step in the context of their specific service. This would help them to demonstrate that their personal information processing is necessary and proportionate. Services must follow the key principles of data protection law such as purpose limitation, data minimisation, and storage limitation. Where providers process special category data (such as biometric face scans to uniquely identify a person), they must apply the additional protection required by data protection law – see the ICO guidance on special category data. See also the ICO guidance on data protection by design and default.

Reduction

4.49 Reduction refers to limiting the circulation and visibility of content rather than removing it entirely.[297] While reduction actions interfere with users' expression less than content removal, content and activity is less visible to users and can pose issues for accountability and oversight of these decisions.[298]

4.50 Reduction is often used by service providers for content that is deemed by a service provider to be misleading, offensive, or otherwise risky, but not illegal. Services which allow user-to-user sharing of content mediated by recommender algorithms and search services which present users with search results in response to user queries may find it useful to introduce reduction techniques.

4.51 This step could include:

a) **Designing recommender systems** that promote content diversity and variety, which might include content featuring diverse perspectives. For example, recommender systems can be designed to ensure users encounter a variety of content, which can help mitigate the risk of rabbit holes (i.e. feedback effects which funnel users towards content of increasing thematic intensity) and prevent the formation of echo chambers/filter bubbles (i.e. thematically homogenous content or content with a unified theme).
b) **Reducing the prominence of misogynistic abuse and sexual violence from content recommender feeds** to minimise widespread dissemination, and the risk of users organically encountering it in large volumes.[299] Algorithmic impact assessments and other algorithmic evaluations (see Action 2) can also help providers understand how recommendation algorithms are operating, and how they can best address the risks of harm from the promotion of violent, abusive and hateful content.
c) **Removing links** to sites dedicated to hosting non-consensual images, or to services such as nudification apps used to generate non-consensual intimate content, as well as removing user-generated content on nudification sites, which amounts to an advertisement for these sites.
d) **De-monetising** user-generated content which promotes misogynistic abuse and sexual violence in breach of the provider's terms of service to prevent it from earning advertising income.[300]
e) **Reducing the prominence of misogynistic abuse and sexual violence in search results,** for example by downranking such content or promoting content that is high-quality, yet relevant to the search query (see **Case study 10**).
f) **Scanning for duplicates**: When someone successfully requests the removal of explicit non-consensual content featuring them from search, making sure to scan for – and remove from search results – any duplicates of that image that are found.
g) **Blurring nudity and harmful content** using automated detection, giving users the option to unblur it if they want to see it.

---

[297] Gillespie, T., 2022. Do Not Recommend? Reduction as a Form of Content Moderation, *Social Media + Society*, 8 (3). [accessed 31 October 2025].
[298] Gillespie, T., 2022. Do Not Recommend? Reduction as a Form of Content Moderation, *Social Media + Society*, 8 (3). [accessed 31 October 2025].
[299] Appelman, N., 2023. Disparate Content Moderation Mapping Social Justice Organisations Perspectives on Unequal Content Moderation Harms and the EU Platform Policy Debate. [accessed 6 November 2025].
[300] Jankowicz, N., Gomez-O'Keefe, I., Hoffman, L. and Vidal Becker A. 2024. It's Everyone's Problem: Mainstreaming Responses to Technology-Facilitated Gender-Based Violence. [accessed 6 November 2025].

h) **Imposing rate limits** (such as a limit of how many comments or posts a user can make in a specific time period) to prevent mass-posting in pile-ons.

i) **Sharing information** regarding what kinds of posts might trigger downranking, de-prioritisation, or exclusion for transparency. Providers should carefully consider the benefits and risks of notifying individual users when their posts are deprioritised or excluded: on the one hand, such notifications improve transparency and accountability. On the other hand, they can help malicious users 'game' the system to bypass safety measures.

4.52 **Case study 13** sets out examples of how a general search service can reduce the accessibility of websites and forums that host intimate image abuse content.

Case study 13 (good practice): Gender-sensitive search services

Scenario

After being shown evidence about the rising use of nudification apps from third-party partners, the provider of a **general search service** aims to reduce the accessibility of websites and forums that host **intimate image abuse content**, including nudification apps and sexualised deepfakes.

Steps to take

The provider decides to introduce a range of different interventions to reduce access to intimate image abuse content. It knows that no single intervention will be able to tackle all the content, so it focuses on putting in place multiple layers of protection. The provider:

- Delists such websites and URLs, preventing identified intimate image abuse content from appearing in search results.
- Downranks URLs with harmful content to deprioritise it in search results and reduce its visibility.
- Signposts to its reporting systems when a user enters a relevant query by providing clear, accessible tools for users to report and request removal of intimate image abuse content.
- Further evaluates the impacts of these interventions on an ongoing basis, continuing to monitor content appearing in search results to identify emerging risks and adapt interventions accordingly.

Considerations

- Providers should monitor for unintended consequences, such as suppression of URLs which do not need to be restricted under the Act or policies set by the provider about how search content will be actioned.
- Providers should also update policies and safety measures to respond to evolving harms.

4.53 **Case study 14** sets out examples of how a social media provider can use automated detection to detect and respond to digital misogynoir.

Case study 14 (good practice): Automated detection of misogynoir content and results

Scenario

A **social media service provider** is provided with evidence by a frontline organisation that it is failing to detect digital misogynoir, the intersectional form of online abuse affecting Black women that combines racism and misogynistic abuse.[301]

Existing automated moderation tools are missing content specific to this intersectional harm.[302] Abusive language or terminology aimed at Black women can include specific abusive terms or language[303], meaning there can be gaps in a detection algorithm if it is not programmed to recognise these words in an abusive context. The provider needs to improve its detection systems to better identify and address this form of abuse.

Steps to take

The team responsible for automated detection tools is alerted by the trust and safety team to this harm. It investigates, identifies key weaknesses in the current detection tools, and sets out to improve them:

- It trains content detection models on datasets that reflect intersectional abuse, including self-reported experiences of misogynoir so it detects incidences with more accuracy.
- It reviews the services' content labelling processes so labelling data reflects differences in how language may be used in different communities, particularly those at risk of harm.
- It further collaborates with researchers and advocacy organisations to refine detection methods and ensure they reflect lived experiences (see **Case study 3**) as well as using trauma-informed surveys to gather views from its user base affected by this harm to inform future changes (see **Case study 6**).

Considerations

- The provider should regularly evaluate the impact of changes to ensure they are reducing harm without introducing new biases.
- The provider should regularly evaluate its detection tools to address the risk of both over-blocking (for example, blocking content recounting a survivor's experience of online-gender violence) and under-blocking (for example, missing intentional misspellings of slurs or coded allusions to violence).
- The provider should have regard to the needs of their UK user base in considering what languages are needed to ensure its automated systems are accurate at detecting harmful content.
- When using content moderation tools, providers must also consider privacy rights and comply with data protection law requirements.[304]

---

[301] Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive. [accessed 29 October 2025]; Bailey, M., 2021. Misogynoir Transformed: Black Women's Digital Resistance. [accessed 28 October 2025].

[302] Evidence has found that existing automated detection tools are ineffective at detecting misogynoir due to a lack of sensitivity to context. This issue is 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. See Kwarteng, J., 2022. Misogynoir: Challenges in Detecting Intersectional Hate, Social Network Analysis and Mining, 12 (1). [accessed 20 October 2025].

[303] Glitch, 2023. The Digital Misogynoir Report: Online abuse against Black women allowed and enabled to thrive.[accessed 29 October 2025]

[304] See the ICO's Guidance on Content moderation and data protection for further information.

5. Providing support

Overview

Context

5.1 This chapter considers how service providers can embed safety-by-design to effectively support users and address harm, including once a product is in operation. This includes user control tools that give users more agency over their experiences, user support tools that provide helpful and accessible information, and reporting systems that support survivors and victims when harm occurs.

5.2 Some services may have very basic user controls, user support, or reporting mechanisms. Other services may have more developed mechanisms, but they may not sufficiently account for online gender-based harms. For example, many services do not offer users easily navigable tools or clear information about how to use them.[305] A Refuge study found that 95% of survivors who reported online abuse to a service were not satisfied with the response they received – and over half did not receive a response at all.[306]

5.3 In this chapter, we set out actions service providers can take to improve women and girls' online safety through user controls, user support, and reporting mechanisms. The specific actions include:

- **Action 7**: Give users better control over their experiences.
- **Action 8:** Enable users who experience online gender-based harms to make reports.
- **Action 9:** Respond appropriately when online gender-based harms occur.

5.4 For each action, we set out our expectation of what a baseline of safety looks like (**'foundational steps'**) for service providers to meet their duties to protect UK users. We have a range of enforcement powers to hold companies to account where they fail to comply with their duties.

5.5 We also highlight additional **good practice steps** to illustrate how providers can build on the foundational steps to create safer experiences for women and girls, give their users more autonomy, and provide assurance that users can seek appropriate redress for any harm that does occur.

5.6 Harm response tools are especially important as some perpetrators of online gender-based harms are highly motivated to cause harm and will attempt to evade safety measures – even if the measures have been designed and deployed to prevent abusability.

5.7 Service providers should not put an undue burden on users to be fully responsible for their own safety online. The user controls, user support and reporting tools set out in this chapter should be supported by strong groundwork from the provider to ensure that responsibility for user safety does not sit solely with users of the service.

[305] W3C, 2024. Web Content Accessibility Guidelines (WCAG) 2.2. [accessed 16 November 2025].
[306] Refuge, 2022. Marked as Unsafe.[accessed 16 November 2025].

Our target outcomes

5.8 Women and girls should have more agency to shape their online lives, and those that experience harm should be offered appropriate support. Providers should offer user control, user support, and reporting tools that account for the dynamics and complexities of online gender-based harms.

5.9 If providers develop and deploy user controls and user support, women and girls will gain more control over the content and users they encounter online and will have accessible information available to them to make informed choices about risk. This will allow them to make personalised decisions about what safety looks like for them – including the ability to make changes if their circumstances change.

5.10 Appropriate harm response also includes providers designing and operating reporting systems which are easy to find, easy to use, and fit for purpose, as well as supporting survivors and victims when harm occurs. This includes specialised support for online gender-based harms such as stalking and coercive control.

5.11 Media literacy also plays an important role in preventing online gender-based harms by equipping users with the skills and critical thinking needed to engage effectively with user control, user support, and reporting tools. Tools to give users more control and facilitate reporting should be designed in line with the Best Practice Design Principles for Media Literacy.

Action 7: Give users better control over their experiences

5.12 Users often feel that they lack control over their online experiences.[307] Women and girls frequently experience unwanted behaviour and encounter unwanted content on online services.[308] Too often, providers do not deploy easily accessible and navigable tools, nor do they provide users with clear information about how to change their content and safety settings.[309] This makes it harder for users to curate their experiences online and increases the risk of harm.[310]

5.13 Providers can empower users by providing them with greater and more granular control over their own experiences. Increased options over who contacts them, what they see and what information about them is visible can allow users to reduce the risk of experiencing or encountering gender-based harms and minimise their impact when they occur.[311]

---

[307] Centre for Data Ethics and Innovation, 2020. Online targeting: Final report and recommendations. [accessed 30 October 2025].
[308] Ofcom, 2024. Experiences of using online services.
[309] Service providers should consider their obligations under other relevant legislation (for example, the Equality Act 2010), as well as industry standards and good practice to ensure their services meet the access needs of disabled users. See also W3C, 2024. Web Content Accessibility Guidelines (WCAG) 2.2. [accessed 10 January 2025].
[310] Centre for Data Ethics and Innovation, 2020. Online targeting: Final report and recommendations. [accessed 30 October 2025]; The Behavioural Insights Team, 2020. Active Online Choices: Designing to Empower Users. [accessed 30 October 2025].
[311] Grimani, A., Gavine, A. and Moncur, W., 2022. An evidence synthesis of covert online strategies regarding intimate partner violence, *Trauma, Violence, & Abuse*, 23 (2). [accessed 30 October 2025].

Increased control enables users to provide feedback on the content they see and make informed choices about what safety looks like for them.

5.14 Providers can use tools such as choice bundles (see Chapter 4) to make it easier for users to engage with increased options over their online experiences.[312]

Foundational steps: What are the expectations for service providers?

5.15 Our Codes set out the following steps for service providers based on functionality, risk and size:[313]

a) **Block and mute:** User-to-user service providers should offer users the options to block[314] and mute[315] other users' accounts.[316] (see **Case study 15**).
b) **Disable comments**: User-to-user service providers should offer users a feature which allows them to disable comments on their own posts.[317]
c) **Negative feedback:** User-to-user service providers should offer children a feature which enables them to give negative feedback on content that is recommended to them by a content recommender system.[318]
d) **Group chats:** User-to-user service providers should provide children with the option to accept or decline an invitation to a group chat.[319]
e) **Supportive information:** User-to-user service providers should provide children with supportive information to help them make informed choices about risk. The provider should provide children with supportive information when they restrict interactions with other user accounts or content.[320]
f) **Support materials**: Service providers should provide children with age-appropriate user support materials explaining the user safety tools available to them on a service.[321] The

[312] The Behavioural Insights Team, 2021. Active Online Choices: Designing to Empower Users. [accessed 30 October 2025].
[313] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[314] A user tool where: the blocked user(s) cannot send direct messages to the blocking user; the blocking user will not encounter any content posted by blocked user(s) on the service and vice versa; the blocking user and blocked user(s) if they were connected, will no longer be connected.
[315] A user tool where: the muting user will not encounter any content posted by muted user(s) on the service; and the muted user(s) is not aware that they have been muted and continues to encounter content posted by the muting user.
[316] Illegal Content (ICU J1), Protection of Children (PCU J1).
[317] Illegal Content (ICU J2), Protection of Children (PCU J2).
[318] Protection of Children (PCU E3).
[319] Protection of Children (PCU J3). This measure also appears as a foundational step for 'Action 5: Set safer default settings' as the measure is relevant both for preventing harm and providing support.
[320] Illegal Content (ICU F2). This measure also appears as a foundational step for 'Action 5: Set safer default settings' as the measure is relevant both for preventing harm and providing support. Protection of Children (PCU F2).
[321] Protection of Children (PCU F1, PCS F4).

services should signpost children to support when they report content, or when they post, re-post, or search for harmful content on a user-to-user service.[322]

Good practice steps: How can service providers go further?

5.16 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps to offer users greater and more granular control over their experiences. This can encourage users who may be at heightened risk of harm – for example, public figures or those experiencing stalking or coercive control – to curate what safety looks like for them.

5.17 These tools, in tandem with the tools aimed at preventing and reducing abuse discussed in Chapter 4, could help tackle the 'silencing' of women and girls on online services where their ability or willingness to participate in discourse is constrained by fear of abusive responses. The good practice steps providers could take include:

a) **Visibility settings:** Allowing users to delete or change the visibility settings of content they upload, including content uploaded in the past.
b) **Block and mute:** Providing users with tools to block and mute multiple accounts simultaneously (see **Case study 15**).
c) **Identity verification:** Allowing users to filter out content from all users who have not completed identity verification.
d) **User controls:** Providing adult users with greater control over what content is recommended to them by content recommender systems.
e) **User preferences:** Allowing users to signal what kind of content they do not want to see, and what kind of content they want to see more of (see **Case study 16**).
f) **Supportive information:** Signposting users to information including explanatory resources on the most suitable user or privacy options for their needs or to support when they provide negative feedback on content they have encountered. Supportive information could also include resources from frontline support organisations on gender-based harms and information about where and how to report a crime.[323]

[322] Protection of Children (PCU F3, PCU F4, PCU F5). This measure also appears as a foundational step for 'Action 6: Reduce the circulation of content depicting, promoting or encouraging online gender-based harms' as the measure is relevant both for preventing harm and providing support.
[323] Department for Science, Innovation & Technology and PUBLIC, 2025. Platform Design and the Risk of Online Violence Against Women and Girls. [accessed 29 October 2025].

5.18 **Case study 15** sets out examples of how a service provider can support women facing pile-ons and coordinated harassment by improving blocking and muting settings.

Case study 15 (good practice): Customisable blocking and muting options

Scenario

A **social media service provider** becomes aware through comments made in the media by women in the public eye that they are facing difficulty preventing **pile-ons and coordinated harassment** on their service. This is resulting in those women refraining from posting on or using the service. Some report receiving hundreds of threatening, abusive and hateful comments, reposts and tags from different accounts which they are unable to prevent. Many of the posts use identical threatening language.

Steps to take

- The service already allows users to disable comments on a post during the pile-on to hide any existing comments and prevent any further comments. The social media service provider goes further to provide users with more flexible options to block and mute. The service:
    - provides 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.
    - allows users to block any current or future accounts connected to a particular phone number or email address.
    - offers users more automated blocking options. For instance, if a user sees a post that is offensive or disturbing to them, they are given the option to block or mute not only the post's author but all users who have reposted during a pile-on.
- The service also adds a rate-limiting feature, preventing accounts from making hundreds of posts within a matter of seconds. This prevents accounts from being able to send abusive messages in quick succession with no deterrence.

Providing users with additional options reduces the burden of safety work on survivors and victims and enables them to reduce their risk of experiencing pile-ons and coordinated harassment.

Considerations

- User tools are part of a wider picture of actions needed by users to prevent pile-ons and coordinated harassment. If any posts received in a pile-on amount to illegal harassment, abuse, or hate, the service provider has requirements under the Act to remove this content (see Chapter 2).
- Ofcom's research found that female politicians felt simpler and more effective reporting systems were crucial alongside user control tools to prevent online abuse.[324] [325]

[324] Ofcom, 2025, Experiences of online hate and abuse among women in politics.

[325] Under the user empowerment duties, Category 1 service providers will also have to offer adult users the option to prevent non-verified users from interacting with their content and to reduce the likelihood of encountering content which non-verified users generate, upload or share.

5.19 **Case study 16** sets out examples of how a discussion forum can use content filters to enable users to opt out of seeing misogynistic abuse online.

Case study 16 (good practice): Content filtering

Scenario

After undertaking user research, a **discussion forum** discovers that many users feel they do not have control over the content in their feeds. Feedback from the users indicates they feel exposed to harmful and unwanted content when they use the forum.

Multiple users give the example of a recent viral trend for posting content that involved some users posting online **misogynistic abuse**, and how they had no way to opt out of seeing this content, besides blocking users who posted it. The discussion forum considers how best to give users more autonomy over their experience on their service.

Steps to take

- The discussion forum adds an option for users to use content filters.[326] Content filtering tools 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.
- Users can apply a filter for the hashtags and keywords specific to the trend they want to avoid. They then stop seeing posts related to the trend in their feeds.
- The tool enables users to shape their online experience and avoid content which contains words and topics likely to be offensive, disturbing, or upsetting to them.

Considerations

- Filters do not replace service providers' duties to remove content that meets the threshold for illegal content, including illegal threats, abuse, and hate (see Chapter 2).[327]

[326] Ofcom research suggests content filters could make users feel more in control of their social media experience. It also suggests that the way in which service providers communicate the outcome of using content control tools to users could be important and rigorous testing of features should be undertaken to understand this. User Empowerment - Technical report

[327] Under the user empowerment duties, Category 1 service providers will also have to offer their adult users features that they can use to reduce their likelihood of encountering certain categories of content, including legal hate and abuse.

Action 8: Enable users who experience online gender-based harms to make reports

5.20 Users who experience or encounter online gender-based harms may face challenges in reporting this to providers.[328] Where reporting systems do not effectively account for harms such as stalking, coercive control and image-based sexual abuse, survivors and victims can experience invalidation and re-traumatisation from the failure of providers to recognise the harm they have experienced, as well as frustration caused by lack of action to tackle the harm.[329]

5.21 Poor experiences with reporting systems erode trust with users and many survivors and victims stop reporting online gender-based harms to providers as a result.[330] It is important for providers to recognise that user reporting relies on survivors and victims of online gender-based harms, and that reporting processes are time-intensive and risk re-traumatising survivors and victims.

5.22 Improving reporting systems is crucial, as this will allow users to inform providers when harm occurs so that providers can take further action where appropriate. This includes allowing users to be able to add relevant context to their report where appropriate, particularly for specific harms like stalking and coercive control.[331]

5.23 Providers should also account in their design of reporting systems that systems can be manipulated with mass false reports to target women and girls during a pile-on.[332]

5.24 Providers can encourage and enable users to make reports by designing reporting systems which are accessible, transparent, easy-to-use, and account for the specific dynamics of online gender-based harms. A safety-by-design approach incentivises providers to encourage and enable users to report online gender-based harms. For instance, they can design trauma-informed reporting systems which consider the specific needs and requirements of survivors and victims.[333]

5.25 Where a user engages with a service provider as part of the reporting process, the user could exercise their data protection rights under UK GDPR at the same time.[334]

---

[328] Department for Science, Innovation & Technology and PUBLIC, 2025. Platform Design and the Risk of Online Violence Against Women and Girls. [accessed 29 October 2025].
[329] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025]; The Open University (Jurasz, O.), 2024. Online violence against women: a Four Nations study. [accessed 30 October 2025].
[330] Victims Commissioner (Storry, M. and Poppleton, S.), 2022. The Impact of Online Abuse: Hearing the Victims' Voice. [accessed 30 October 2025]; PEN America (Vilk, V. and Lo, K.), 2023. Shouting into the Void. [accessed 30 October 2025].
[331] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025]; Flynn, A., Powell, A., Scott, A. and Cama, E. [accessed 10 August 2025].
[332] International Center for Journalists (Posetti, J. and Shabbir, N.), 2022. The Chilling: A global study of online violence against women journalists. [accessed 1 October 2025]
[333] Chayn (Hussain, H.), 2021. Chayn's trauma-informed design principles. [accessed 10 November 2025].
[334] We encourage providers to consult the ICO guide to individual rights.

Foundational steps: What are the expectations for service providers?

5.26 Our Codes set out the following steps for service providers, based on functionality, risk and size:[335]

a) **Complaints processes**: Service providers' complaint processes should enable users, affected persons, and interested persons to make relevant complaints[336] (see **Case study 17** and **Case study 20**).
b) **Complaints systems:** Service providers' complaint systems are easy to find, easy to access, and easy to use and allow users to add supporting information to their complaints.[337]
c) **Complaints communications**: Service providers' complaint systems acknowledge receipt of complaints, provide indicative timeframes,[338] and set out information about how the complaint will be handled[339] (including giving users the option to opt out of communications from a service).[340]
d) **Predictive search**: Search service providers should give users means to easily report predictive search suggestions, and those which the provider determines present a clear and material risk of users encountering illegal content or content harmful to children should no longer be presented to users.[341]

[335] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[336] Illegal Content (ICU D1, ICS D1), Protection of Children (PCU D1, PCS D1).
[337] Illegal Content (ICU D2, ICS D2), Protection of Children (PCU D2, PCS D2).
[338] Illegal Content (ICU D4, ICS D3), Protection of Children (PCU D4, PCS D4).
[339] Illegal Content (ICU D5, ICS D4), Protection of Children (PCU D5, PCS D5).
[340] Illegal Content (ICU D6, ICS D5), Protection of Children (PCU D6, PCS D6).
[341] Illegal Content (ICS F1), Protection of Children (PCS F1).

5.27 **Case study 17** sets out examples of how a pornography service provider can enable a non-user to report intimate image abuse on its service.

Case study 17 (foundational): Reporting as a non-user

Scenario

A **pornography service provider** becomes aware through a third-party organisation that survivors and victims of **intimate image abuse** cannot contact them to get an image taken down without signing up to their service.

Alongside the victim depicted in the image, there may also be others who are affected by the intimate image abuse and face the same barriers as a non-user. This could include a parent or adult with responsibility for a child, or an adult providing assistance to another adult using the service.

Steps to take

- The provider enables its reporting system to allow **external reports** of intimate image abuse, without a user having to go through the account creation process.
- The victim of intimate image abuse then also has the option to ask another individual to report images on their behalf. This can reduce the risk of re-traumatisation.

Considerations

- The service provider could go further to incorporate the good practice step on hash matching for intimate image abuse (see Action 6). This would mean the image would be detected and taken down automatically, saving further effort and distress for survivors and victims if it were to be re-uploaded.
- It could also take down or hide images reported through dedicated intimate image reporting channels immediately, while the report awaits review by moderators.
- It could also partner with a frontline organisation with expertise on intimate image abuse to provide further specialised support to survivors and victims (see **Case study 21** on trusted flaggers).

Good practice steps: How can service providers go further?

5.28 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps to further strengthen their reporting systems and create more trauma-informed and tailored process for reporting online gender-based harms. This could encourage more users to report content and help users to provide the information a service needs to make an appropriate decision. This would lead to better experiences for women and girls, as well as giving service providers a better picture of how harms are manifesting on their service.

5.29 The good practice steps providers could take include:

a) **Exit buttons:** Providing a 'quick exit button' throughout the reporting process which immediately takes the user out of the reporting system. This can protect the privacy of users at a higher risk of abuse.[342]
b) **Report tracking:** Allowing users to track and manage their reports and tailor their experience throughout the complaints process. This could include the option for users to share their reports more easily so they are able to share their own report information with third parties, including frontline support organisations or law enforcement, if they wish (see **Case study 19** and **Case study 20**).
c) **User feedback:** Allowing users to give feedback to the service provider on their reporting process.[343]
d) **Trusted flaggers:** Establishing a trusted flagger programme in partnership with organisations that have expertise in online gender-based harms.[344] (see **Case study 21**).
e) **Incident reporting:** Allowing users to report incidents of abuse with the appropriate context, such as being able to include multiple posts or interactions in one report, including abuse that happened on another service or offline. Service providers may take various approaches in responding to these reports (see **Case study 22**).[345]
f) **Media literacy:** Adopting the principles of user-centric design and timely interventions for supportive information in Ofcom's Best-Practice Design Principles for Media Literacy.
g) **Support during reporting:** Signposting to relevant supportive materials when reports are made on gender-based harms, such as stalking and coercive control or image-based sexual abuse. These materials could include information about frontline support organisations and information on how to report a crime, where appropriate.[346]

[342] For further information on how to develop exit buttons and how they work see: Exit a page quickly – GOV.UK Design System [accessed 25th August 2025].
[343] Service providers should have regard to the needs of their UK user base in considering what languages are needed when designing their reporting process.
[344] Refuge response to the 2023 Illegal Harms Consultation, p.12; Suzy Lamplugh response to the 2023 Illegal Harms Consultation, p.21; SWGfL response to the 2023 Illegal Harms Consultation, p.15; UCL Gender and Tech response to the 2023 Illegal Harms Consultation, p.10; VAWG Sector Experts response to the 2024 Protection of Children Consultation, p.13; Impact of social media and screen-use on young people's health - Science and Technology Committee - House of Commons
[345] Service providers should have regard to the needs of their UK user base in considering what languages are needed when designing their off-service incident reporting process.
[346] Service providers should have regard to the needs of their UK user base in considering what languages are needed for supportive information.

5.30 **Case study 18** sets out examples of how a gaming provider can support users being harassed by improving its reporting options.

Case study 18 (good practice): Reporting options

Scenario

A **gaming service provider** becomes aware through content posted by women gamers online that they are receiving **misogynistic abuse** on the service. The gamers report that they face barriers reporting this abuse, because it occurs through in-game audio chats and virtual reality spaces, and the service's reporting system only allows reports about written chats.

Harassment does not always occur through written and visual communication such as images, comments, and direct messages which can be easily recorded and reported.

Steps to take

The gaming service provider offers **accessible reporting** for all types of content and interaction supported on the service. This prevents perpetrators from avoiding the provider's enforcement systems by using non-written or visual communication. The provider:

- Allows the user to report the abuse that happened during an in-game voice chat.
- Allows the user to report the perpetrator sexually harassing them and invading their physical space on its virtual reality platform.
- To keep on top of perpetrator tactics, the provider updates its reporting systems when it makes any changes to its service, including new user interaction features.

Considerations

Adding additional reporting options could increase the volume and type of reports received. Providers should consider appropriate resourcing and training for moderation teams so they are prepared to respond to this.

5.31 **Case study 19** sets out examples of how a social media service provider can improve user trust on reporting systems by being more transparent.

Case study 19 (good practice): Transparent reporting process

Scenario

A **social media service provider** becomes aware through user testing that confidence in its reporting system is low.

Survivors and victims report they feel underinformed about the providers' process of assessing the report, the actions it could take, who will be informed and how long the process will take.[347] This can be particularly distressing when a victim is reporting a **repeat perpetrator**, as they may not know if reporting will result in the perpetrator being informed of the report.[348]

Steps to take

A team at the service provider is tasked with improving the reporting process. To do this, it draws on user feedback and external research to introduce the following changes with **transparency** in mind. The provider:

- Adds realistic estimated time periods for responding to reports where possible, so the user knows when to expect a response.[349] This is particularly important for high-risk user reports.
- Includes information about the possible enforcement action outcomes that correspond to different policy violations during the reporting process, so the user clearly understands what could happen, who could be informed about the report and how a reported user may be sanctioned.
- Offers users feedback options during multiple stages of the reporting process. The user feedback is used to inform the future updates to the reporting system, ensuring survivors and victims' voices are heard and the system is improved in the future.

Considerations

- An accessible and transparent reporting process can build trust and better manage expectations between users and providers.[350] As reporting processes are likely to involve processing of personal data, ensuring that providers are transparent about how user data is used can also support providers to comply with their requirements under data protection law.[351]
- Providers should balance transparency in reporting with awareness of perpetrator behaviour, ensuring information about its reporting or sanctions processes is not used by perpetrators or bad actors to evade enforcement action (see **Case study 7** on abusability product testing).
- User feedback surveys should be trauma-informed (see **Case study 6**).

---

[347] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025].
[348] Responses to our February 2025 consultation: Refuge.
[349] Chayn (Hussain, H.), 2021. Chayn's trauma-informed design principles. [accessed 30 October 2025].
[350] World Wide Web Foundation, 2021. Tech Policy Design Lab: Online Gender-Based Violence and Abuse. [accessed 2 August 2025].
[351] We encourage providers to consult the ICO's UK GDPR guidance and resources: ICO's UK GDPR guidance and resources, in particular the ICO guidance on the right to be informed and transparency.

5.32 **Case study 20** sets out examples of how a social media service provider can ensure that its reporting process is effective for survivors and victims of coercive control.

Case study 20 (good practice): Tracking and managing reports

Scenario

A frontline domestic abuse organisation alerts a **social media service provider** that survivors and victims of **coercive control** are facing issues with tracking reports they've made to the service. Experiences of online gender-based harms are often complex and highly contextual and frequently involve multiple interactions or pieces of content. Reporting is also time-consuming and can be re-traumatising for survivors and victims.[352]

Survivors and victims can face barriers in being able to share their reports with third parties.[353]

Steps to take

The provider decides to provide a **reports dashboard** to its users.

- The provider displays all of a user's previous reports in one place. The user can then check the status of their reports easily, instead of having to contact the provider about the status of each report individually.
- The provider adds the functionality to download or share copies of the reports. This allows a user to share information with a third party, such as a frontline support service or law enforcement.
- The provider also gives the user the option to invite a trusted contact to support them with the reporting process.
- Providing greater choice over the reporting process can provide survivors and victims of coercive control with greater agency and transparency over the process and build trust between the user and the provider.[354]

Considerations

- The Web Foundation's Tech Policy Design Lab developed a prototype for a reporting dashboard that enables users to view their reports and see when reports are resolved, with consideration of avoiding re-traumatisation.[355] The dashboard could also allow users to add additional context to their reports and collect and archive evidence of harmful content.
- Functionalities supporting a user to download or share a report should be compliant with data privacy laws. Providers should consult ICO guidance.[356]

[352] Chayn (Hussain, H.), 2021. Chayn's trauma-informed design principles. [accessed 30 October 2025]. Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025]
[353] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025]
[354] PEN America (Vilk, V. and Lo, K.), 2023. Shouting into the Void. [accessed 30 October 2025].
[355] World Wide Web Foundation, 2021. Tech Policy Design Lab: Online Gender-Based Violence and Abuse. [accessed 30 October 2025].
[356] Users have the right to obtain access to their personal information under data protection law. See the ICO guidance on subject access requests for further information. For more information see ICO's UK GDPR guidance and resources. [accessed 5 September 2025]

5.33 **Case study 21** sets out examples of how an image sharing service provider can establish a trusted flagger programme to tackle intimate image abuse.

Case study 21 (good practice): Trusted flaggers

Scenario

An **image sharing service provider** is alerted by a frontline organisation with expertise in **intimate image abuse** that the service is being used by perpetrators to share intimate image abuse content, and the organisation is facing barriers getting the images taken down.

Having found out intimate images of themselves have been posted online, a survivor or victim of intimate image abuse might first seek help from a trusted frontline organisation to help them navigate the situation, rather than approaching the provider first.

Steps to take

The provider establishes a **trusted flagger programme**, where it partners with the frontline intimate image abuse organisation. Trusted flagger programmes can assist survivors and victims by building relationships between service providers and organisations with expertise in harms, such as stalking and coercive control or intimate image abuse.[357]

- The trusted flagger can raise cases to the provider on behalf of survivors and victims and the provider can ensure the cases are prioritised, lowering the chance of re-traumatisation and ensuring content is taken down swiftly.
- The trusted flagger can provide additional support to the survivor and victim throughout their experience.
- The trusted flagger is also able to use this relationship to alert the provider to emerging forms of harm, or new ways the service is being abused by perpetrators.

Considerations

- To ensure effectiveness, providers should 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.
- Trusted flaggers should be treated fairly as partners and provided with necessary resource and support.[358]

[357] Ofcom Stakeholder Workshop 2 on Women and Girls Online Safety, 19 November 2024.

[358] 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 emphasises that trusted flagger organisations should be provided with the necessary resource and support when carrying out additional work to make a service safer.

5.34 **Case study 22** sets out examples of how a livestreaming service can support survivors and victims of stalking by introducing a policy on reporting harmful off-service behaviour.

Case study 22 (good practice): Reporting off-service behaviour

Scenario

A **livestreaming service provider** receives reports from creators using the service that perpetrators of **stalking** can harass them through its commenting and messaging functionalities on livestreams.

Harms such as stalking are characterised by a wider pattern of behaviour, both online and offline.

Steps to take

Building upon its existing policies, the service provider introduces an additional policy that enables users to report harmful **off-service behaviour** (see **Case study 1** on capturing stalking in terms of service). This allows users of the service who have experienced stalking offline to inform the provider.

- The provider evaluates the report and, if satisfied that sufficient verifiable evidence has been given, the provider takes enforcement action.
- Enforcement action includes 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.

Considerations

- When handling and evaluating off-service reports, providers should comply with data protection law where this involves processing of personal data. [359] Services should take a data protection by design approach and ensure that necessary safeguards are integrated to protect the rights and freedoms of users.[360] Providers should ensure fairness in data processing, particularly if automated systems are used to analyse reports or determine enforcement actions. Any use of technology for this purpose must be transparent and free from bias. [361]

[359] Providers should consider ICO Guidance on data protection principles. Depending on the nature of the data reported, service providers may be processing criminal offence data. Where this is the case services should consult the ICO guidance on criminal offence data.
[360] See the ICO guidance on data protection by design and default.
[361] Providers should consult the ICO Guidance on automated decision-making and profiling and AI and data protection.

Action 9: Respond appropriately when online gender-based harms occur

5.35 When providers do not allocate sufficient resource and expertise to ensure appropriate action is taken when users experience or encounter online gender-based harms on their services, it can cause further distress to survivors and victims. With insufficient resource and expertise, providers may fail to respond to user reports, and their content and search moderation systems may not identify harmful content. In other cases, providers do act but without specific consideration of the nuances of gender-based harms and the importance of survivor and victim agency.[362]

5.36 Online gender-based harms are often highly contextual. Harms such as intimate image abuse, stalking or coercive control may not immediately be identified by content or search moderation. Existing evidence shows particularly poor response rates for reports of intimate image abuse.[363]

5.37 Providers can reduce the impact of online gender-based harms by taking appropriate action when it occurs on their service. Content and search moderation that accounts for harms such as intimate image abuse, stalking and coercive control enables providers to identify patterns of harmful behaviour. Improving responses to user reports allows providers to better support women and girls who have experienced online gender-based harms.

Foundational steps: What are the expectations for service providers?

5.38 Our Codes set out the following steps for service providers based on functionality, risk and size:[364]

a) **Taking action**: User-to-user service providers should have a content moderation function that allows for the swift take down of identified illegal content and swift action against identified content harmful to children.[365]

[362] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025]; Refuge response to the 2023 Illegal Harms Consultation, p.4; Glitch response to the 2023 Illegal Harms Consultation, p.6.
[363] Qiwei, L., Zhang, S., Kasper, A., Ashkinaze, J., Eaton, A., Schoenebeck, S. and Gilbert, E., 2024. Reporting Non-Consensual Intimate Media: An Audit Study of Deepfakes. [accessed 30 October 2025].
[364] The 'foundational steps' refer to a range of expectations we have already set out for service providers at the time of publication of this guidance. The 'foundational steps' apply to different services based on functionality, risk, and size. Information about which foundational step applies to which service is set out in the Guidance at a Glance document. The Illegal Content Codes of Practice apply to all regulated user-to-user and search services, and the Protection of Children Codes of Practice apply to such services where they are likely to be accessed by children (within the meaning of the Act).
[365] Illegal Content (ICU C2), Protection of Children (PCU C2).

b) **Performance targets**: Service providers should set and record targets for content and search moderation functions.[366]
c) **Prioritisation**: Service providers should prepare and apply policies on the prioritisation of content for review.[367]
d) **Moderation teams**: Service providers should secure that content and search moderation functions are resourced[368] and individuals working in moderation receive training and materials (see Action 1).[369]
e) **Complaints**: Service providers should handle complaints about suspected illegal content and content harmful to children in accordance with content prioritisation processes.[370]
f) **Appeals**: Service providers should determine appeals promptly or according to performance targets,[371] depending on the provider. Complainants should, as far as appropriate and possible, be restored to the position they would have been in had the decision not been made, following upheld appeals.[372]

Good practice steps: How can service providers go further?

5.39 In this section, we set out good practice steps that service providers can take in tandem with the foundational steps to further embed understanding of online gender-based harms into their systems and processes. This will enable them to respond to harm in a way that supports survivors and victims, while also minimising the risk of future harmful behaviour and content circulating on their services.

5.40 The good practice steps providers could take include:

a) **Enforcement action:** Taking consistent and clearly communicated action against users who violate a service's terms of service (see **Case study 24**). To maximise effectiveness, services should, where possible, take steps to identify and prevent the creation of new accounts by perpetrators who may do so to evade a sanction or a ban.[373]
b) **Fact-checking and labelling:** Adding fact-checking and labelling to content can be a useful tool to address gendered disinformation, which can be used in coordinated harassment campaigns against women and girls in public life.[374] [375]

---

[366] Illegal Content (ICU C4, ICS C3), Protection of Children (PCU C4, PCS C4).
[367] Illegal Content (ICU C5, ICS C4), Protection of Children (PCU C5, PCS C5).
[368] Illegal Content (ICU C6, ICS C5), Protection of Children (PCU C6, PCS C6).
[369] Illegal Content (ICU C7, ICS C6), Protection of Children (PCU C7, PCS C7).
[370] Illegal Content (ICU D7, ICS D6), Protection of Children (PCU D7, PCS D7).
[371] Illegal Content (ICU D8, ICS D7, ICU D9, ICS D8), Protection of Children (PCU D8, PCS D8, PCU D9, PCS D9)
[372] Illegal Content (ICU D10, ICS D9), Protection of Children (PCU D10, PCS D10).
[373] At the time of publication, Ofcom's consultation on Additional Safety Measures has closed, and we are carefully considering responses. It contains additional proposals for platforms to take greater action on user bans and sanctions. It includes the proposal to '*ban users who share, generate, or upload CSEA, and those who receive CSAM, and take steps to prevent their return to the service for the duration of the ban'*. It also proposes that *'providers should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share illegal content/content harmful to children and/or illegal content/content harmful to children proxy, with the objective of preventing future dissemination of illegal content'*. Consultation: Online Safety - Additional Safety Measures - Ofcom.
[374] Internet Governance Forum, 2021. Best Practice Forum on Gender and Digital Rights: Exploring the concept of gendered disinformation. [accessed 30 October 2025]; National Democratic Institute, 2021. Addressing Online Misogyny and Gender Disinformation: A How-To Guide. [accessed 30 October 2025].
[375] Providers should consider the type of content and context when deciding on how to present AI-generated information. Labelling AI-generated media online | PNAS Nexus | Oxford Academic [accessed 15 August 2025]

c) **Watermarks and metadata:** Adding watermarks and provenance metadata to AI-generated content can show where and how it has been created or edited. This could help survivors and victims tracing the origin of deepfake intimate image abuse content.[376]
d) **Moderator review:** Sending high risk and highly contextual user reports of online gender-based harms for review by specifically trained moderators[377] (see **Case study 23**).
e) **Hiding content:** Hiding potentially harmful content while it is assessed in content moderation, such as potential intimate image abuse, can minimise harm to survivors and victims while the assessment is completed.
f) **Reporting channels:** Creating dedicated reporting and review channels for online gender-based harms.[378]

[376] Attribution measures that indicate whether content is synthetic on their own won't address the harm posed by deepfake intimate image abuse, but can help in tracing the origins of an image back to its source. Deepfake Defences 2 - The Attribution Toolkit, Ofcom 2025.
[377] Automated content moderation tools can be effective. However, it is important that service providers recognise that certain user reports of online gender-based violence are nuanced and highly contextual and so human moderators with specific training are likely to be highly effective in addressing these.
[378] Service providers should have regard to the needs of their UK user base in considering what languages are needed when designing their dedicated reporting and review channels for online gender-based harms.

5.41 **Case study 23** sets out examples of how a social media service provider can train its content moderation team to improve its response to coercive control behaviour.

Case study 23 (good practice): Coercive control training

Scenario

A moderation team at a **social media service provider** responsible for assessing user reports of **coercive control** occurring through its messaging and posting features does not have specific awareness of this harm.[379] Through a feedback mechanism on its reporting system, it identifies increasing frustration from survivors and victims that their reports are not being evaluated correctly, and no action is being taken.

Coercive control is perpetrated in complex and highly personal ways, and online coercive control often replicates and extends the same dynamics as offline coercive control.[380] When reviewing reports, the team misses that the abuse is contextual and may be a harmful pattern of behaviour rather than a single interaction or piece of content.[381]

Steps to take

The service provider provides specific training to its content moderation teams on coercive control. This enables moderators to better identify instances of this harm and respond to user reports.

- The trained moderation teams now take into account considerations such as how to respond to user reports appropriately without escalating offline risks to survivors and victims.
- The teams further their understanding of coercive control and how it occurs on the service through partnerships with organisations which have frontline experience and expertise (see **Case study 21** on trusted flaggers).

Considerations

- Content moderators need to receive adequate support from providers to undergo training and carry out this work.
- As highlighted in Action 8, it is crucial that survivors and victims of coercive control are able to give adequate context to their reports.

[379] Woodlock, D., McKenzie, M., Western, D. and Harris, B., 2020. Technology as a Weapon in Domestic Violence: Responding to Digital Coercive Control, *Australian Social Work,* 73 (3). [accessed 14 March 2025].
[380] Woodlock, D., McKenzie, M., Western, D. and Harris, B., 2020. Technology as a Weapon in Domestic Violence: Responding to Digital Coercive Control, *Australian Social Work,* 73 (3). [accessed 14 March 2025].
[381] Refuge, 2021. Unsocial Spaces. [accessed 5 September 2025].

5.42 **Case study 24** sets out examples of how a messaging service provider can improve its response to perpetrators who continually harass women and girls on its service.

Case study 24 (good practice): Action on repeat perpetrators

Scenario

A **social media service** offers a GenAI feature that can be added into a chat between users. The feature can be used to generate text that users can send to one another.

Its moderation team notices the same perpetrators being reported by many users and identifies a pattern of behaviour towards women and girls, where the perpetrators are **harassing** them with large volumes of sexually explicit content they generate using the GenAI feature.[382]

A small number of users are often responsible for a large amount of online gender-based harm, particularly in cases of coordinated harassment.[383]

Steps to take

The provider introduces a **strike-based enforcement system**, where a user receives a 'strike' against their account for misuse of the service.

- When a user attempts to generate harmful content through the GenAI feature through sexually explicit prompts, they receive a strike against their account.
- If a user receives multiple strikes, their access to the GenAI feature is removed for a given period of time. If a perpetrator continues to misuse the service, repeated strikes lead to an account ban.
- When a user receives a strike, they are informed *why they are receiving it and what the consequences are*. The service also includes the functionality for users to appeal strikes and related enforcement action.
- This prevents perpetrators from getting to the stage where they can share harmful content by intervening early. Applied effectively, warning-based enforcement systems can act as a form of deterrence preventing a single act of abusive behaviour from becoming a pattern.[384]

Considerations

- 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 service providers to consult guidance from the ICO, which applies to all forms and methods of moderation.[385]
- Service providers may wish to take into consideration whether the account belongs to a child or an adult where appropriate and relevant to the nature of the offence.

---

[382] Wilson Center (Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S. and Kaufmann, Z.), 2021. Malign Creativity: How gender, sex, and lies are weaponized against women online. [accessed 29 December 2025].

[383] Ofcom Stakeholder Workshop 2 on Women and Girls Online Safety, 19 November 2024.

[384] A 2021 study found that posting warning messages to users believed to be at risk of suspension can reduce the use of hateful language on some social media networks. Yildirim, M.M. et al. (2023) 'Short of Suspension: How Suspension Warnings Can Reduce Hate Speech on Twitter', Perspectives on Politics, 21(2), pp. 651–663.

[385] See ICO guidance Content moderation and data protection and Profiling tools for online safety.

A1. Glossary

A1.1 This glossary sets out definitions of terms used throughout the Guidance.

Term	Definition
Abusability testing/evaluations	Abusability testing or abusability evaluations draw on the concept of "usability" which is used to evaluate how easy it is for users to navigate a website or device to accomplish their goals. In contrast, abusability evaluations test how easy it is to abuse a tool or feature for harm, and therefore point to ways that abusability can be minimised in design.
Action	Actions are intended to assist user-to-user and search service providers in designing, testing, deploying and operating their services in a manner that takes responsibility for the safety of women and girls on their services.
Blocking	To take action that will result in the blocking user and blocked user being unable to send direct messages to each other or encounter each other's content, and to become unconnected if they were connected.[386]
Chilling effect	Describes how individuals suppress self-expression due to fears of legal penalties or social backlash, even without direct threats.
Choice architecture	The way that choices (such as rankings or defaults) are presented to individuals. Choice architecture can have a significant impact on the choice users make.[387]
Coercive control	This covers the offence of controlling or coercive behaviour in an intimate or family relationship.[388] For the purposes of helping service providers understand the sort of content and activity likely to be included, this is a kind of repeated or continuous behaviour often associated with domestic abuse.
Connection list	A feature that allows users to find other users that they may not already be connected with. This list can sometimes be suggested by the platform based on user interests and existing connections.
Consent	Agreement by choice, where a person has the freedom and capacity to make that choice.

[386] A more precise definition is contained in paragraph ICU J1.3 of Recommendation ICU J1 (user blocking and muting) in the Illegal content Codes of Practice for user-to-user services.

[387] The Competition and Markets Authority (CMA) has described a taxonomy of 21 online choice architecture practices that influence consumers through the way in which choices are presented (such as defaults and ranking), the information that is presented (such as framing and complexity of language), and the pressure applied to consumers' choices (such as reminders and scarcity claims). See CMA, 2022, Online Choice Architecture, How digital design can harm competition and consumers [accessed 19 October 2025].

[388] An offence under section 76 of the Serious Crime Act 2015 (controlling or coercive behaviour in an intimate or family relationship). For more information on the risks associated with controlling or coercive behaviour, see section 5 of the Illegal Harms Register of Risks.

Term	Definition
Co-opt	In the context of this document, we use this term to refer to using a good or service in a way that is different from the usual or intended purpose.
Content recommender systems	An algorithmic system which determines the relative ranking of an identified pool of content (that includes regulated user-generated content) from multiple users on content feeds. Content is recommended based on factors that it is programmed to account for, such as popularity of content, characteristics of a user, or predicted engagement. References to content recommender systems do not include a content recommender system employed exclusively in the operation of a search functionality which suggests content to users in direct response to a search query, product recommender systems or network recommender systems.
Counterspeech	The practice of responding to speech that seems harmful or offensive. It can take many forms such as challenging, debunking or critiquing harmful speech, amplifying alternative viewpoints, providing accurate information, and fostering empathy and understanding.
Cyberflashing	This refers to the offence of sending an image of genitals, with intent to causing alarm, distress or humiliation, or to obtain sexual gratification with recklessness as to whether alarm, distress or humiliation is caused. [389] The sending of unsolicited sexual images without the consent of the recipient can amount to cyberflashing.
Deepfake	Forms of audio-visual content that have been generated or manipulated using AI, which misrepresent someone or something. It is an offence and a form of intimate image abuse to create, or request the creation of, deepfake intimate images without the consent of the individual depicted.[390]
Doxing	Non-consensual public disclosure of private or sensitive information about an individual, such as their location, phone number or email address.[391] Doxing can take place in the context of coordinated harassment or coercive control.

[389] The definition of the criminal offence of cyberflashing is contained in section A4 in the Illegal Content Judgements Guidance.

[390] An offence under section 66E-66H of the Sexual Offences Act 2003.

[391] The Crown Prosecution Service provides a similar definition, see: CPS, 2024. Cybercrime – prosecution guidance. [accessed 10 November 2025].

Term	Definition
Duties	The legal duties we are required to focus on for the purposes of this Guidance under section 54 of the Act are set out in Parts 3 and 4 of the Act and are applicable to providers of Part 3 services (user-to-user and search services). The Act imposes duties which require providers to identify, mitigate and manage the risks of harm from illegal content and activity, and content and activity that is harmful to children. Certain additional duties are set out in Part 4 of the Act, including some which apply only to categorised services.
Foundational step(s)	Expectations we have already set out for service providers relevant to achieving an action. This includes actions service providers can take to help them comply with their duties related to risk assessments and transparency, and when looking to implement measures set out in the Codes of Practice, in the context of protecting women and girls.
Gaslighting	Psychological manipulation of a person usually over an extended period of time that causes the victim to question the validity of their own thoughts, perception of reality or memories, and typically leads to confusion, loss of confidence and self-esteem, uncertainty of one's emotional or mental stability, and a dependency on the perpetrator.
Gendered disinformation	Gendered abuse online that uses misleading or false gender-based narratives against women and their participation in public life. It is a combination of online disinformation through falsity and coded language to evade moderation, detection and coordination.
Generative AI or GenAI	Refers broadly to machine learning models that can create new content. Models create a wide variety of outputs including text, images, video, and audio.
Good practice steps	Further information on how services can tackle online gender-based harms that builds on the expectations we have set out in the foundational steps.
Harm prevention	Refers to systems which attempt to anticipate and mitigate risk before harm happens. In the context of product development, it can include testing to identify potential routes for abuse, and to allow for changes in features and functionalities to prevent harm.
Harm response	Refers to systems which address harm and minimise its effects, leading to support or restitution for the person who experienced harm. In the context of online gender-based harms users experience or encounter on a service, this includes easy to find and to use reporting systems which are fit for purpose, and taking appropriate action to address the impact of harm.

Term	Definition
Hash matching	Hash matching is an umbrella term for techniques that create a 'fingerprint' of a given piece of content. In practice this means using an algorithm to analyse content and create a 'hash' that can represent it. Hashes are then stored in a database that can be accessed by multiple parties as required. In the context of online safety, online platforms can use hashing to notify other platforms of illegal or harmful content they have identified, and vice versa. Hashing databases exist for CSAM, terror content, and non-consensual intimate images.
Image-based sexual abuse	For the purposes of this Guidance, this includes intimate image abuse (see definition), self-generated indecent imagery (see definition) and cyberflashing (see definition).
Intimate image abuse	This includes all the offences of intimate image abuse, which differ slightly across the UK.[392] These offences are about sharing or threatening to share intimate images without the consent of the person depicted. Other terms for this include "non-consensual intimate image abuse" (commonly referred to as "NCII") and "revenge porn".
Intersectionality	The interconnected nature of social categorisations such as race, class, and gender, regarded as creating overlapping and interdependent systems of discrimination or disadvantage; a theoretical approach based on such a premise. The term was coined in 1989 by Kimberlé Crenshaw describing how traditional feminist theory and antiracist policies exclude Black women who face overlapping discrimination unique to them.[393]
Malign creativity	The use of coded language, iterative and context-based visual and textual memes and other tactics to avoid detection on social media platforms.
Media literacy	The ability to use, understand and create media and communications across multiple formats and services. Ofcom has specific media literacy duties as set out in Chapter 2 of this document.
Misogynistic abuse and sexual violence	Content and activity which normalises, encourages, or reinforces attitudes and behaviours which promote abuse, violence or hatred targeted at women and girls.[394]

[392] The definition of the criminal offence of intimate image abuse is contained in Section A11 of the Illegal Content Judgements Guidance.

[393] Crenshaw, K., 1989. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics, *University of Chicago Legal Forum*, 1989(1). [Accessed 29 October 2025].

[394] Misogynistic abuse and sexual violence can cover the criminal offences of threats, abuse and harassment (including hate) as defined in section A4 in the Illegal Content Judgements Guidance. It can also cover the following kinds of content harmful content defined in the Guidance on content harmful to children: pornographic content (section 2), abuse and hate content (section 6) and violent content (section 8).

Term	Definition
Monitoring and assurance	Function to provide independent assurance that measures taken to mitigate and manage the risks of harm to individuals identified in the risk assessment are effective on an ongoing basis. This function should report to, and its findings should be considered by, either: a) the body that is responsible for overall governance and strategic direction of a service; or b) an audit committee.
Muting	To take action that will result in the muting user not encountering the content of the muted user unless the muting user visits the user profile of the muted user.[395]
Non-designated content	A category of content harmful to children defined in the Act as content, which is not primary priority content or priority content, of a kind which presents a material risk of significant harm to an appreciable number of children in the United Kingdom.[396]
Nudges	Design features which lead or encourage users to follow the designer's preferred paths in the user's decision making.
Online gender-based harms	Harmful and abusive content and activity that disproportionately affects women and girls. For the purposes of the Guidance, we focus on four overlapping forms of harm: misogynistic abuse and sexual violence, pile-ons and coordinated harassment, stalking and coercive control, and image-based sexual abuse.
Perpetrator	A user, individual or group who conducts and participates in online gender-based harms.
Pile-ons and coordinated harassment	Content and activity that involves many perpetrators targeting an individual victim or small group of victims with abusive, hateful or threatening content, often repeatedly or at scale.[397]
Product	An all-encompassing term that includes any functionality, feature, tool, or policy that a service provides to enable users to interact with or use the service.
Primary priority content	A category of content that is harmful to children, as defined in section 61 of the Act.[398] This includes pornography, suicide, self-harm, and eating disorder content.
Priority content	A category of content that is harmful to children, as defined in section 62 of the Act.[399] This includes hate and abusive speech.

[395] A more precise definition is contained in paragraph ICU J1.3 of Recommendation ICU J1 (user blocking and muting) in the Illegal content Codes of Practice for user-to-user services.
[396] Section 60(2)(c) of the Act.
[397] Pile-ons and coordinated harassment can cover the criminal offences of threats, abuse and harassment (including hate) as defined in section A4 in the Illegal Content Judgements Guidance. It can also cover the following kinds of content harmful content defined in the Guidance on content harmful to children: abuse and hate content (section 6) and violent content (section 8).
[398] We have typically grouped the different kinds of primary priority content as follows: pornographic content, suicide and self-harm content, eating disorder content.
[399] We have typically grouped the different kinds of priority content as follows: abuse and hate content, bullying content, violent content, harmful substances content, dangerous stunts and challenges content.

Term	Definition
Red teaming	A type of evaluation method that seeks to find vulnerabilities in GenAI models.
Repeat perpetrators	A user, individual or group who repeatedly conducts and participates in online gender-based harms. In many cases, the majority of online gender-based harms comes from a relatively small proportion of a service's user base that is highly motivated to continue this pattern of behaviour.
Reduction	Limiting the circulation and visibility of misleading or harmful content rather than removing it entirely.
Re-traumatisation	The re-experiencing of thoughts, feelings or sensations experienced at the time of a traumatic event or circumstance in a person's past. This can be triggered by, for example, encountering a smell or a sound associated with the traumatic event, or by being put into a similar situation or experience.
Safety-by-design	A proactive approach to integrating safety considerations into the design and development of products, systems, or processes.
Safety work	Online and offline strategies women employ to respond to, avoid, and cope with gender-based violence. This can include avoiding certain actions (like posting online or walking alone at night), spending time thinking or planning about safety risks, or moderating how they dress or present themselves online to avoid experiencing violence. Safety work is an unfair burden which limits women's space for action and self-expression, and makes them responsible for preventing violence.
Search service	An internet service that is, or includes, a search engine. For the purpose of this Guidance, we primarily focus on "general search services" which are services that enable users to search the internet by inputting search requests. A general search service derives search results from an underlying search index developed by either the provider of the service or a third party. Search results are presented using algorithms that rank based on relevance to a search request (among other factors). This is because we don't have substantial evidence on the risks of online gender-based harms on vertical search services.[400]
Self-generated indecent imagery	Child sexual abuse material created by the child depicted in the image.[401]
Semen images	Images or videos where semen is depicted on top of a non-intimate image.

[400] For more information on vertical search services, please see: Protecting people from illegal harms online - Annex 3: Glossary.

[401] For more information on self-generated indecent imagery, see paragraphs 2.27-2.28 in the Illegal Harms Register of Risk. The definition of the criminal offence of child sexual exploitation and abuse is contained in sections A5 and A6 of the Illegal Content Judgements Guidance.

Term	Definition
Service provider	We refer to a platform as 'service', and the legal entity that provides the service as 'provider' or 'service provider'. In other words, a website might be the 'service', and the company that owns and runs it would be the 'provider'.
Sextortion	For the purposes of this Guidance, this refers to a form of intimate image abuse that involves sharing, or threatening to share, intimate images. This may be to extort money or force the victim to do something against their will. Images are often taken or made without the victim realising or consenting.
Supportive information	Also referred to as Crisis Prevention Information. Refers to information provided by a search service in search results that typically contains the contact details of helplines and/or hotlines, and links to trustworthy and supportive information provided freely by a reputable and reliable organisation.
Survivor and victim	A person who has experienced online gender-based harms.
Stalking	This covers the offences of stalking, versions of which exist across the UK.[402] For the purposes of helping service providers understand the sort of content and activity likely to be included, this is a form of harassment, characterised by a pattern of fixated, obsessive, unwanted, and repeated behaviour which is intrusive.
Trauma-informed approach	An approach which acknowledges (i) the prevalence of trauma; (ii) how trauma affects all individuals involved with the programme, organisation, or system, including its own workforce; (iii) and responds by putting this knowledge into practice. This often includes an emphasis on informed consent as well as content warnings. Examples include trauma-informed design and reporting systems.
Usability	The quality or state of being convenient and practicable for use. In the context of product design, it is a measure of how well a specific user in a specific context can use a product/design to achieve a defined goal effectively, efficiently and satisfactorily.
User centric design	Refers to a design process which is inherently iterative and puts user needs at the centre of every stage of this process, to create highly usable and inclusive products.
User-to-user service	An internet service by means of which content that is generated directly on the service by a user of the service, or uploaded to or shared on the service by a user of the service, may be encountered by another user, or other users, of the service.
User controls and user support	Refers to tools which give users more agency over the content, users, and experiences they encounter online and accessible information to make informed choices about risk.

[402] The definition of the criminal offence of stalking is contained in section A4 in the Illegal Content Judgements Guidance.

Term	Definition
Victim-blaming	Explicitly stating or implying that the victim is to blame for the abuse they have experienced. It often focuses on actions that a victim could have taken (or not taken) to avoid experiencing abuse.
Viral	A piece of content – such as a post, image, or video – that achieves a high level of popularity by being quickly and widely shared online, particularly on social media.
Virality	The degree to which online content spreads easily and/or quickly across many online users, alongside how much engagement and/or views a piece of content received (such as 'shares', 'likes', and 'view', etc.).