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

Consultation response form

Your response

Question	Your response
Question 1: Do you have any comments on our proposed approach to 'content and activity' which 'disproportionately affects women and girls'?	Confidential? – N The Free Speech Union (FSU) recognises that women and girls can face specific and serious risks online, including stalking, coercive control, and intimate image abuse. Addressing these harms, where they involve clearly unlawful conduct, is both legitimate and necessary. However, we have significant concerns about the conceptual and regulatory framework proposed in this guidance, particularly in relation to content and activity that is lawful but deemed to "disproportionately affect" women and girls. Ofcom's framing of this broader mission is itself cause for concern. In the consultation preamble, it outlines the statutory duties imposed on platforms to tackle illegal content and content harmful to children. It then adds that it is "also" required to produce guidance on how providers "can take action" against "harmful content and activity that disproportionately affects women and girls". Elsewhere, it states – twice – that technology firms "should do more" to improve women and girls' online safety. Such language blurs the boundary between unlawful content and lawful expression that is simply considered undesirable by the regulator. Combined with the repeated assertion that firms "should do more", this framing imposes a moral expectation that goes beyond the law, exerting pressure on providers to pre-emptively restrict lawful content. The emphasis on "safety by design" (2.23) reinforces this trajectory, suggesting that platforms should build their systems in ways that suppress certain types of speech from the outset, even where no legal obligation to do so exists. This conceptual slippage carries through into the guidance itself, which encourages platforms to go "above and beyond" their legal duties under the Online Safety

Question	Your response
	Act (Section 1.11), and to "go further" in addressing gender-based harms (Section 2.46). What Ofcom refers to as an "ambitious vision" (p.4) signals a broader shift in the role of regulation: from the enforcement of law to the performance of virtue. The guidance urges platforms to adopt a series of "good practice" recommendations that are explicitly non-binding but framed as moral obligations. These extend well beyond foundational legal duties and include measures such as algorithmic downranking of "harmful" content and behavioural nudges to deter users from posting material deemed inappropriate. Ofcom also states that it will assess and report on the extent to which platforms have implemented these measures (Section 2.87), effectively placing public pressure on those that do not. In this context, the label "voluntary" begins to resemble de facto obligation, driven by reputational and moral pressure rather than legal duty. In borrowing its logic from corporate managerialism, this "above and beyond" approach conflates legal responsibility with moral leadership. In the private sector, incentivising individuals to exceed performance targets may bring financial reward. But in the realm of regulation, exceeding legal thresholds is not just unnecessary – it risks undermining core legal protections, such as the duty on providers to have a particular regard for free speech. Regulation must operate within clear legal limits, not be guided by shifting cultural expectations or reputational incentives. Ofcom's justification for this shift rests on some dubious psychological assumptions, as well as vague, unscientific language. "Misogynistic speech is often not illegal," it acknowledges, "but, at scale, it can normalise harmful beliefs in boys and men and impact women and girls' experience both online and offline." (p.3) It further claims that social media algorithms are "rewarding" such content with greater reach, and urges platforms to act accordingly. But what exactly is "misogynistic speech" and who gets to define it? While few would deny that misogyny exists as a social problem, the term is both broad and subjective, stretching from the unambiguously hateful to the merely provocative or politically unfashionable. Would Ofcom class Jordan Peterson's online posts as "misogynistic", for instance? What's more, Ofcom's own decidedly behaviourist framing suggests that speech does not have

Question	Your response
	to be overtly misogynistic in intent to be problematic. Instead, it is enough that, "at scale", it might shape attitudes in undesirable ways. The potential consequences of such an elastic framework are difficult to overstate. Ofcom does not define "misogynistic speech" with any legal precision, yet it implies that content may be subject to moderation not because it is illegal or even overtly abusive, but simply because it might, at scale, contribute to harmful cultural attitudes, where "harmful" seems to mean undesirable beliefs, rather than actual, physical harm to women and girls. This creates a risk that platforms will suppress not only high-profile provocateurs, but also more ambiguous forms of expression. Laddish or immature 'banter' by teenage boys, for instance, may well be crude and distasteful, but it is also often part of a broader developmental process through which young people test boundaries, encounter social resistance, and learn what is and is not acceptable. Even if one wished to regulate such content, automated moderation systems are blunt tools. They cannot reliably distinguish between sincere belief, mockery, satire, ironic reversal, or performative repetition. A single phrase might function as hate speech in one context, self-deprecating humour in another, or parody in a third. These complexities of language-in-use – so central to the expressive life of a democratic society – are precisely what algorithmic systems struggle to parse. Yet it is this very ambiguity, this polysemy, that makes online life dynamic, pluralistic, and worth defending. The public square is not made healthier by suppressing every expostulation that might offend. It is made healthier by allowing ideas, even misguided or immature ones, to be confronted through argument, satire, humour, or dissent. We are dubious about the behaviourist assumptions underlying Ofcom's approach. The idea that "misogynistic speech", even when lawful, can be treated as a kind of external stimulus that shapes attitudes "at scale" presumes a passive audience and a stable, one-directional relationship between language and behaviour. But speech does not function in this way. Meaning is not transmitted like a signal; it is negotiated in context, and its effects are shaped by irony, intention, reception, and resistance. In this light, the idea that harmful attitudes can be neutralised – or even changed – by removing content overlooks the active role users play in interpreting,

Question	Your response
	challenging, and often mocking the very ideas that platforms are being encouraged to suppress. The law, properly understood, is not a ladder to be climbed in pursuit of moral purity, but a boundary designed to balance competing rights in a democratic society. When suppression of content goes beyond the law in one area, it may often fall foul of the law in another – Article 10 of the European Convention on Human Rights (ECHR) being the obvious example here. The outcome of Ofcom's guidance may be a quieter and more curated internet, but it will also be a less intellectually generous one. It will be less capable of accommodating disagreement and the messy friction through which social understanding evolves.
Question 2: Do you have any comments on the nine proposed actions? Please provide evidence to support your answer.	Confidential? – N The FSU's response to Question 2 addresses only those actions that, in our view, pose the most significant risks to freedom of expression. We focus in particular on Actions 1, 3, 4, 5, and 6. We have not commented on the remaining actions, which we consider either less relevant to the protection of lawful speech or outside the scope of our organisational remit. *Action 1: Governance and Accountability* Ofcom recommends that governance processes "address online gender-based harm" through measures such as consulting subject matter experts and setting internal policies. In principle, the FSU supports efforts to ensure that platform governance structures are responsive to genuine safety concerns, particularly where those risks relate to coercive control or unlawful abuse. However, we are concerned by the absence of any reference to expertise on freedom of expression in these recommendations. The use of the term "gender" rather than "sex" also suggests that the "harm" Ofcom is seeking to prevent would include content that trans or non-binary people – or their allies – find offensive. For instance, does Ofcom believe JK Rowling's tweet of her smoking a cigar to celebrate the recent Supreme Court judgment on the Equality Act is a species of "gender-based harm" and should be suppressed? Would a gender-critical feminist "misgendering" a trans person be classed as a "gender-based harm"? We would be more comfortable if Ofcom used the phrase "sex-based harm" throughout, making it clear it does not intend to side with trans activists over gender-critical feminists when it comes to encouraging platforms to suppress content.

Question	Your response
	If platforms are to design policies or set moderation rules in response to "gender-based harms" – especially where that harm includes lawful but controversial speech – then protections for freedom of expression must be embedded at the governance level. Article 10 of the ECHR, as incorporated into UK law via the Human Rights Act 1998, requires that lawful political expression, particularly on matters of public interest, be given the highest level of protection. As was noted in *R v Casserly* [2024] EWCA Crim 25, the greater the value of the speech in question, the weightier must be the justification for interference. That obligation cannot be discharged if speech rights are treated as a secondary consideration once internal safety protocols are already in place. This concern is particularly acute in the context of gender-critical speech, which is both politically contested and legally protected. The consultation document defines misogyny in broad and non-statutory terms, including "content that actively encourages or reinforces misogynistic ideas or behaviours" and content that "normalises… harmful sexual behaviour" (2.8). It is not clear whether this includes speech that rejects gender-identity-based definitions of womanhood and affirms biological sex as the basis for women's rights. If it does, then Ofcom may be encouraging platforms to treat gender-critical views as misogynistic, despite the fact that such views are protected under UK law. This raises two fundamental questions for Ofcom. First, where does Ofcom stand on the definition of "woman" for the purposes of assessing misogyny? If womanhood is treated as a self-declared gender identity, does that mean gender-critical speech – such as objecting to male-born individuals in women's spaces, or stating that only biological females are women or refusing to use the preferred gender pronouns of trans people – will be treated as "misogynistic"? If so, Ofcom risks promoting guidance that conflicts with the UK Supreme Court's judgment in *For Women Scotland Ltd v Lord Advocate* [2022] UKSC 31, which confirmed that "woman" in UK law must be interpreted according to biological sex. Second, even if Ofcom does not take a position on the definition of "woman", or prefers to remain officially neutral, what steps will it take when US-based Big Tech firms adopt self-ID-based moderation policies and begin to shadow-ban, suppress, or categorise gender-critical speech as "harmful", only to claim that they are complying with Ofcom's guidance? This is not a hypothetical

Question	Your response
	concern. We already see platforms suppressing lawful gender-critical speech under internal hate speech policies – for years, Twitter penalised users for "misgendering" – and Ofcom's ambiguous framing of "misogyny" provides no assurance that such overreach will be challenged or corrected. Unless Ofcom is willing to state clearly and publicly that gender-critical beliefs – which are lawful, protected and recognised by UK courts – must not be treated as a form of misogyny, this guidance risks enabling ideological censorship, with lawful speech suppressed by moderation regimes that do not reflect UK law or jurisprudence. Political expression, including debates about sex-based rights, falls within the core of what Article 10 of the ECHR is intended to protect. That protection is particularly robust where the views expressed are controversial or challenge dominant cultural narratives. Article 10 protects not only well received expression, but also that which offends, shocks or disturbs (*Kwiecien v Poland* (2007) 48 EHRR 7 at [43]) If Ofcom's governance recommendations result in policies or enforcement practices that chill such speech – whether through direct domestic influence or indirect alignment with global platform standards – it will have failed in its obligation to uphold the balance of rights at the heart of democratic discourse. Ofcom should consider the principle of proportionality in *Bank Mellat v HM Treasury* (No 2) [2013] UKSC 39: (1) whether the objective of the measure is sufficiently important to justify the limitation of a protected right; (2) whether the measure is rationally connected to the objective; (3) whether a less intrusive measure could have been used without unacceptably compromising the achievement of the objective; and (4) whether, balancing the severity of the measure's effects on the rights of the persons to whom it applies against the importance of the objective, to the extent that the measure will contribute to its achievement, the former outweighs the latter. We would urge Ofcom to make clear that platform governance processes must balance safety with robust protections for lawful expression, including beliefs and view-

Question	Your response
	points that may be politically contested or culturally sensitive. This means not only consulting safety experts, but also engaging legal and civil society voices to ensure that freedom of expression is not sidelined in favour of reputational risk management or institutional orthodoxy. *Action 4: Abusability Evaluations and Red-Teaming* Ofcom's guidance recommends that providers "conduct abusability evaluations and product testing", including through "red teaming to identify ways malicious actors may try to use service features to perpetrate harm" (Section 2.58). We recognise that red-teaming can be a valuable tool for identifying tangible safety risks, particularly where platform features might be exploited to facilitate stalking, coercive control, or the non-consensual sharing of images. However, as currently framed, red-teaming focuses solely on how features might be abused in ways that cause harm. It does not consider the equal and opposite risk: that measures designed to prevent abuse may themselves result in the suppression of lawful speech. If red-teaming is to play a meaningful role in platform design, it should be applied symmetrically. Ofcom should encourage providers to test not only how features might be exploited to cause harm, but also how those same features, or the mechanisms introduced to mitigate their misuse, might chill democratic discourse. For example: • Could a content filter disproportionately suppress protected beliefs, such as gender-critical views? • Could a default prompt steer users away from important journalism or minority political opinions? • Could a bundled "safe mode" setting make it effectively impossible to encounter dissenting views on gender, race, or religion? If Ofcom is serious about empowering users and protecting rights, it should also advocate for red-teaming exercises that anticipate over-removal, not just under-removal. That means asking platforms to assess how new features or safety systems might inadvertently undermine freedom of expression, interfere with political speech, or disproportionately impact lawful but culturally contested perspectives. *Actions 3 and 5: Transparency and Safer Defaults*

Question	Your response
	We consider these two actions together because both are concerned not with the content of moderation rules, but with the ways platforms structure user experience and influence behaviour through design. It is in these system-level interventions – defaults, prompts, and interface framing – that the most serious risks to lawful expression may arise. What appears to be a neutral, technical matter of user interface design is, in practice, being framed by Ofcom as a mechanism for politically charged content moderation. This direction is evident in the consultation's call for platforms to track the "effectiveness of safety measures" (Action 3) and to "set safer defaults" (Action 5), with the stated aim of helping services prevent harm "through the design of their services" (Table 2, p.19). The document adds: "Our research shows that default settings can be a powerful tool to encourage safer behaviour online" (2.61). That statement is footnoted to Ofcom's 2024 trial report *Behavioural Insights to Empower Social Media Users*, produced in partnership with the Behavioural Insights Team (BIT). We have reviewed that research, which tests whether users could be steered away from so-called "sensitive content", and we find its implications deeply concerning from a freedom of expression standpoint. The report confirms that user engagement with content controls was significantly influenced by changes to defaults, salience, and timing. Users exposed to heightened salience cues were more likely to opt for content filtering. When default options were pre-selected, even fewer users changed their settings. Most (88%) did not revisit their initial choice at all, illustrating just how "sticky" these settings can be. Crucially, the content categories tested in Ofcom's trial included "hate speech" and "misinformation" – i.e., terms that have no statutory definition in UK law, and which remain vague, subjective, and ideologically capacious. Yet these are precisely the kinds of categories that platforms may now be encouraged to filter, as part of user control systems designed to reduce exposure to gender-based harms. In this light, Ofcom's behavioural research does not simply map user preferences – it explores how to shape those preferences using interface design and emotional timing. These behavioural ambitions are also evident in Ofcom's report *Understanding Online Choices, Preferences, and Welfare* (2024), which states that "intervention is also more likely to be justified where negative externalities

Question	Your response
	arise from certain categories of content", citing as justification content that might "harm democracy" or constitute "viral misinformation". Again, these terms – just like Ofcom's definition of misogyny – are not defined in law, and are conceptually nebulous and vulnerable to selective or politicised interpretation. By grounding regulatory intervention in such expansive categories, Ofcom lays the conceptual groundwork for the indirect suppression of lawful speech via behavioural design, rather than open legal restriction. Ofcom's emphasis in the consultation document is overwhelmingly on privacy-related default settings, such as geolocation and interaction controls for users at risk of coercive behaviour. This framing appears repeatedly in the guidance (e.g. pp. 28 – 29), where default design is linked to preventing metadata leaks or reducing stalking risk. The regulator rightly highlights the importance of strong defaults in these contexts. Default protections in areas like geolocation privacy or coercive control can clearly enhance user safety without engaging freedom of expression concerns. We also recognise that, as stated in the consultation document, "bundling settings" may assist vulnerable users by simplifying complex interfaces, particularly where a platform includes many features or frequent updates. This point is reinforced on p.51 (A2.26), where Ofcom notes that "bundling settings for services with many features" can help ensure users "always have the most secure and private options". However, we are concerned that these safety bundles may not be confined to privacy settings alone. If content controls – particularly default off settings for lawful but controversial material – are quietly folded into these bundles, the implications for free expression are significant. So-called "misogynistic" content, while potentially offensive to some users, is often entirely lawful and forms part of broader political, cultural, or social debates. These concerns could hardly be described as hypothetical. While the consultation foregrounds privacy-related settings, the behavioural research underpinning Ofcom's position focuses almost entirely on user responses to "sensitive content", including "hate speech" and "misinformation". These categories are vague, subjective and undefined in UK law. If content-based filtering was not on Ofcom's radar, it is difficult to explain why the regulator has invested so

Question	Your response
	heavily in researching how defaults, prompts, emotional salience and timing can influence users' exposure to "sensitive content". The trials show that subtle adjustments in choice architecture can meaningfully shift user behaviour, with most users sticking with initial settings and relatively few revisiting them later. This disconnect raises a serious question about why design interventions tested on political and cultural content are being promoted in a consultation that largely foregrounds privacy-related settings. We would urge Ofcom to include a clear and explicit commitment that any guidance it issues should not be interpreted by platforms as requiring the use of default or bundled content filters for lawful speech, particularly where that speech forms part of ongoing social or political debate. Without such a safeguard, providers may feel compelled to include ideological content restrictions by default in their safety bundles in order to align with perceived regulatory expectations, leading to the suppression of legitimate expression. *Action 6: Reduce the Circulation of Online Gender-Based Harm* Ofcom's proposed Action 6 recommends that platforms take further steps to reduce the circulation of online gender-based harm. This includes proactive tools for detection, removal and reduction, such as hash-matching, link removal and de-prioritisation. The consultation notes that many of these interventions rely on automated systems and may involve "proactive technology" within the meaning of the Online Safety Act (Sections 2.66–2.72). While the removal of clearly illegal content is a legitimate regulatory objective, we are concerned that the scope of this action – particularly as it relates to automated enforcement, undefined harms and expanding notions of "intimate image abuse" – risks entrenching a model of cultural censorship through technology. We support the removal of illegal content such as non-consensual intimate images, and we recognise the value of proactive tools like hash-matching to address such material swiftly and at scale. Ofcom is right to highlight the particular risks of intimate image abuse on adult content platforms and to showcase existing partnerships such as StopNCII.org, which provide a privacy-respecting and rights-compatible framework for content removal. However, what begins as a legitimate safety objective risks overreaching when the definition of harm itself begins to expand.

Question	Your response
	Our concern lies with the scope of what qualifies as "intimate image abuse" – a concept that, while relatively well-defined in UK law, is already the subject of political proposals to expand its meaning in troubling ways. The Commons' Women and Equalities Committee has recently called for the legal definition of an "intimate image" to be broadened to include "culturally sensitive" content, such as photographs of Muslim women without a hijab or Orthodox Jewish women without wigs. Under this proposed framework, images need not contain nudity or sexual content to be classed as NCII; instead, their "intimacy" would be defined by the subjective cultural or religious norms of the person depicted. This approach would enshrine into law a variable, identity-contingent definition of criminality: the same photograph could be lawful or unlawful depending on the ethnicity, religion, or gender identity of the subject. Such a standard raises profound concerns for the rule of law, equality before the law, and freedom of expression. For instance, would posting a picture of a trans woman that dated back to the time in their life when they identified as male count as "intimate image abuse"? The consequences for journalists, human rights organisations and political campaigners are obvious. A documentary photographer covering anti-hijab protests in Tehran might publish images that are clearly in the public interest – but under the proposed UK model, an equivalent image taken in London could be subject to takedown as a "culturally intimate" violation. This chilling effect on political and social commentary would be a direct consequence of expanding the definition of harm beyond what the law currently requires. This legal and conceptual expansion would be troubling enough on its own. But the risks are compounded by the technical systems Ofcom proposes to enforce it. For instance, automated content moderation systems and Ofcom's suggestion that providers layer together tools such as consent nudging, uploader verification and proactive flagging. As the consultation notes, many of these systems fall under the category of "proactive technology" as defined by the Online Safety Act. While automation may be effective for removing clearly illegal content, it also brings the danger of blunt enforcement, especially where the definition of harm is vague, subjective or culturally contingent. Algorithmic moderation systems, which rely on datasets of previously flagged material, exacerbate this problem.

Question	Your response
	These tools routinely struggle to interpret context, cultural references, nuance or satire. This can lead to the suppression of legitimate speech, the misclassification of benign content and the systematic erasure of minority perspectives. The opaque nature of AI decision-making processes further limits accountability and increases the likelihood of bias or error in what are, effectively, "black box" systems. The chilling effect on online discourse is already observable as users become hesitant to engage with certain topics for fear of being flagged by automated systems, ultimately stifling public debate and narrowing the range of perspectives in circulation. Finally, the convergence of broad legal definitions and automated enforcement becomes even more concerning when you take regulatory mission creep into account, which seems inevitable. We also note that the same Committee report that proposed redefining the concept of an "intimate image" has called for the creation of a UK Online Safety Commissioner, explicitly modelled on Australia's eSafety Commissioner. While the Government has formally rejected this recommendation, the proposal itself reflects a broader direction of policy thinking that is unlikely to disappear. Once introduced into the political conversation, such ideas often re-emerge, particularly when framed as pragmatic solutions to online harm. The Australian experience offers a cautionary tale of how well-meaning safety infrastructure can rapidly expand its mandate to suppress lawful but controversial opinions. Initially established to tackle child sexual abuse material and revenge porn, the Australian eSafety Commissioner's remit has since expanded into areas of lawful but controversial speech – ordering takedowns of gender-critical content, pressuring platforms to suppress so-called "hypermasculine" discourse and pursuing global takedown orders for material already geo-blocked in Australia. These powers, once created, are rarely contained. The Free Speech Union of Australia, a sister organisation, has been involved in numerous tribunal battles with the eSafety Commissioner in defence of freedom of expression. If Ofcom's guidance is to remain consistent with the UK's liberal legal tradition, it must clearly resist this shift toward identity-contingent criminality and avoid endorsing enforcement mechanisms that could replicate the Australian model, where a regulator originally tasked with preventing abuse now imposes informal, ideologically-driven censorship of lawful expression. The power to

Question	Your response
	prevent harm must not become the power to suppress dissent. Although the eSafety-style model has, for now, been set aside by the Government, the underlying impulse it reflects remains present in the broader policy conversation. We would urge Ofcom to reject this expansionist trajectory. Efforts to address intimate image abuse must remain tightly focused on material that is clearly illegal under UK law and must not evolve into a vehicle for culturally contingent censorship. Ofcom should make clear that any future guidance on proactive technologies, automated moderation or content classification will respect the foundational liberal principles of objectivity, universality and legal certainty – and will not endorse or facilitate the suppression of lawful speech under the shifting rubric of "cultural harm".
Question 3: Do you have any comments about the effectiveness, applicability or risks of the good practice steps or associated case studies we have highlighted in Chapter 3, 4 and 5? Are there any additional examples of good practices we should consider? Please provide evidence to support your comment.	Confidential? – N We have addressed the applicability and risks of selected good practice steps – particularly in Actions 3, 4, 5 and 6 – in our response to Question 2.
Question 4: Do you have any feedback on our approach to encouraging providers to follow this guidance, including our proposal to publishing an assessment of how providers are addressing women and girls' safety? Do you have any examples or suggestions of other ways we could encourage providers to take up the 'good practice' recommendations?	Confidential? – N As outlined in our responses to Questions 1 and 2, we are concerned that Ofcom's model of encouraging uptake – through reputational signalling and transparency reporting – risks pressuring platforms to suppress lawful speech. We urge Ofcom to make clear that platforms will not be penalised, formally or informally, for choosing to implement only those measures that are strictly necessary to meet their statutory duties.
Question 5: Do you have any comments on our impact assessment, rights assessment, or equality impact assessment? Please provide any information or evidence in support of your views.	Confidential? – N We have incorporated our rights-based concerns into our answers to Questions 1 and 2, with a particular focus on freedom of expression and the protection of lawful political and belief-based speech.

Question	Your response
Question 6: Do you agree that our draft Guidance is likely to have positive effects on opportunities to use Welsh and treating Welsh no less favourably than English? If you disagree, please explain why, including how you consider the draft Guidance could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.	Confidential? – N No comment.

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