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

Response to Ofcom Consultation: A Safer Life Online for Women and Girls

Mihaela Popa-Wyatt (Senior Lecturer, Philosophy, University of Manchester)

Ofcom's draft guidance is a necessary and commendable first step in addressing the disproportionate harms faced by women and girls online. The stakes are high, given the evolving landscape of online gender-based harms. However, to be effective, the final version must be strengthened in clarity, scope, and accountability. The recommendations I propose here are grounded in my academic research on misogyny, extreme misogyny, and the dynamics of hate and violence in incel communities, as well my policy recommendations on holding platforms accountable via a tax on harmful content.

Question 1: On Content and Activity that Disproportionately Affects Women and Girls

(I) **Definitions**. Ofcom's final guidance should include a more robust and philosophically grounded definition of gender-based harms, starting with clear and precise definitions of *sexism*, *misogyny*, *abuse*, and *gender-based violence*, both online and offline.

Rationale:

A major limitation of Ofcom's draft guidance is the lack of conceptual clarity around key terms such as misogyny and gender-based harms. For regulatory frameworks to function effectively, especially in addressing systemic and intersectional harms, definitions must capture not only the surface manifestations of gender-based harms but also their structural, ideological, and institutional roots.

Definition Proposal:

Online misogyny must not be conceptualised as merely a problem of individual hatred, incivility, or hostile behaviour (as suggested in footnote 18 of the draft guidance). Rather, misogyny should be understood as a system of coercion—i.e. a set of ideological, behavioural, and institutional mechanisms that uphold patriarchal power and male dominance through harassment, punishment, intimidation, exclusion, and violence. This coercive system operates through:

- Policing and punishing women for deviating from traditional gender norms.
- Enforcing submission and vulnerability through economic, social, and physical means.
- Perpetuating structural inequality via institutional practices in law enforcement, media, health, education, and policy.
- Silencing and disciplining "non-conforming" women, especially those in public life, such as feminists, politicians, and whistle-blowers.

Misogyny, in this light, is not reducible to individual psychology motivated by hate, hostility, or malice. It is an ideological apparatus backed by structural power relations and cultural norms that reinforce gender hierarchy and subordination of women. It functions to uphold systemic male advantage by punishing perceived violations of patriarchal norms, often through digitally-mediated acts of violence.

Why This Matters:

Without this conceptual depth, platforms and regulators may fail to recognise the full spectrum of gender-based harms, particularly where abuse does not appear overtly hostile but functions to subordinate, silence, and exclude. Clarity in definitions is essential for:

- Effective risk assessments;
- Designing appropriate mitigation strategies;
- Holding platforms accountable to systemic, not just episodic, patterns of abuse.

Proposed Language for Inclusion in Final Guidance:

> *"Misogyny should be understood as a systemic, structural, and coercive phenomenon, embedded in the cultural and institutional fabric of society. It manifests through both online and offline behaviours, practices, content and activities that aim to police gender roles and uphold patriarchal power. Misogyny includes not only hostile or violent actions, but also acts of intimidation, exclusion, economic coercion, and public shaming intended to enforce the subordination of women and girls."*

Incorporating this enriched definition will align Ofcom's guidance with international human rights frameworks, feminist research, and survivor-informed policy. Definitions matter—not only for legal clarity but to ensure that the regulatory system truly reflects the lived experiences of those it aims to protect.

> ***(II) Scope of Harms***: Ofcom's final guidance should expand its scope of recognised gender-based harms beyond the four core categories currently identified (online misogyny, pile-ons and harassment, online domestic abuse, and image-based sexual abuse). It should adopt a broader, systems-level framework that captures the full range of digital harms that function to silence, intimidate, and exclude women and girls from public and digital life.

Rationale:

While the four core harms identified are important, they do not encompass the full reality of how misogyny operates online. A significant body of research and survivor testimony shows that gender-based harms exist on a continuum, from mundane harms (falling under the category of "legal but harmful", including e.g. disparaging and insulting terms, gender stereotypes, misogynistic attitudes) to extreme forms of misogyny and violence (falling under the category of "illegal harms", including e.g. extremist misogyny driven by male supremacy ideology). In addition to the gender-based harms currently canvassed in Ofcom's draft (e.g., deepfake pornography, gendered disinformation, coordinated online harassment campaigns, doxxing), I propose the final guidance should include harms as *systemic tools of exclusion* designed to punish women's visibility and silence dissent. Furthermore, Ofcom's final guidance should explicitly recognise *misogynistic extremism* as a distinct and high-risk category of online harm. This is not simply a reinforcement of patriarchal norms, but it represents a more dangerous ideological escalation—one that promotes and advocates for the complete subjugation of women and legitimises violence as a means of enforcing male dominance. For example, online content and activity grounded in male supremacist ideology often involves incitement violence, glorifying perpetrators, disenfranchising, intimidating, and threatening women and girls. For illustration, we can categorise extreme misogynistic content in two overlapping groups:

1. *Glorification and Justification of Violence*
    - Celebrating perpetrators of male supremacist violence.
    - Praising manifestos or actions of attackers (e.g., incel-motivated mass shootings).
    - Justifying rape and sexual assault using misogynistic myths (e.g., entitlement to sex, victim-blaming).
    - Sharing violent imagery or videos intended to humiliate, dehumanise, or incite fear among women.
2. *Incitement, Instruction, and Creation of Harm*
    - Direct threats of rape, murder, acid attacks, or other forms of gender-based violence.
    - "How-to" guides teaching users how to doxx, harass, stalk, or digitally abuse women (e.g., creating deepfakes, evading content moderation).
    - Disinformation campaigns against women's rights organisations to incite hatred or delegitimise advocacy.

- Recruitment into male supremacist groups or "manosphere" subcultures promoting dominance over women.

These broader gender-based harms constitute a form of *extreme speech and action* that poses risks not just to the individual safety of women and girls, but they undermine the very principles and values of equality and dignity, democratic participation and public security. Misogynistic extremism does not stop at digital expression. Offline acts that aim to control, coerce, or terrorise women should be viewed as part of a broader continuum of extremist gender-based violence. This includes:

- Physical assaults on women activists or healthcare workers.
- Vandalism or attacks on women's shelters, abortion clinics, or rape crisis centres.
- Revenge porn, doxxing, or deepfake pornography aimed at undermining public figures.
- Hosting or facilitating events (online or offline) that serve as incubators for misogynistic mobilisation.

Policy and Regulatory Implications:

- Ofcom should adopt a specific definition of misogynistic extremism that includes both ideological content and coercive acts intended to reinforce male supremacy and suppress women's autonomy.
- Guidance should direct platforms to identify and act upon patterns of glorification, incitement, recruitment, and coordination associated with male supremacist movements.
- Platforms must be required to assess whether their recommendation systems, community guidelines, and enforcement practices allow permissive environments where misogynistic extremism can thrive.
- Content moderation must move beyond a reactive model to proactively detect, demote, and deplatform networks and narratives promoting misogynistic extremism.

Proposed Definition for Inclusion in Final Guidance:

> *"Misogynistic extremism refers to content and behaviours—both online and offline—that promote, glorify, or incite violence against women and girls with the aim of enforcing patriarchal dominance. It includes male supremacist ideologies that legitimise coercion, exclusion, and harm as means of suppressing women's autonomy and rights. This encompasses celebratory or instructional content, direct threats, disinformation, and recruitment efforts, as well as acts of intimidation, digital abuse, and physical violence."*

Without clear recognition of misogynistic extremism as a specific form of harmful online content, platforms and regulators will be unequipped to tackle its growing influence. Ofcom must ensure that guidance anticipates the ideological drivers of gender-based violence and holds services accountable for content and activity that sustain or incite male supremacist harm. Furthermore, a failure to recognise the link between misogynistic violence as both a systemic and motivational act of terrorism, contributes to the normalisation of misogyny both online and offline.

> *(III) Accountability*: Ofcom should require regulated online services to undertake and publish robust, gender-sensitive risk assessments that specifically address how their design features, algorithmic systems, and content moderation practices contribute to the amplification of misogynistic and male supremacist content. Where such content is facilitated or monetised, platforms must be held publicly accountable through enforceable transparency and remediation obligations.

Rationale:
Misogynistic harms are not accidental by-products, but they are the predictable outcomes of platform systems optimised for virality without responsibility. Social media platforms are not neutral intermediaries. They host, amplify, and monetise misogynistic content—often through opaque algorithms and design features that prioritise engagement over safety. These include frictionless virality, poor reporting tools, and inadequate moderation. In many cases, platforms serve as infrastructure for male supremacist radicalisation, enabling the glorification of rape, incitement to violence, and recruitment into extremist communities. The result is not just symbolic or emotional harm but real-world consequences: fear, trauma, exclusion from public life, and gender-based violence. These harms disproportionately affect women and girls, undermining their right to equal participation in digital and democratic spaces.

Ofcom's final guidance must move beyond setting expectations and toward enforceable standards that ensure meaningful, measurable change. To achieve this, I recommend the following action points:

1. *Mandatory Risk Audits*
   Platforms must be required to conduct regular, independent audits evaluating how their architecture facilitates misogynistic harms and radicalisation pathways.
2. *Design Accountability*
   Platforms should redesign features that enable harm (e.g., default public profiles, algorithmic amplification of hate) and demonstrate that safety-by-design principles are actively implemented.
3. *Transparency and Enforcement*
   Ofcom should enforce public disclosure of gender-disaggregated data on content moderation outcomes, algorithmic impacts, and survivor support efficacy.
4. *Consequences for Non-Compliance*
   Failure to mitigate misogynistic extremism should result in regulatory penalties, public naming in compliance reports, and the potential withdrawal of protections under the Online Safety framework.

Addressing the Structural Nature of Online Misogyny and Gender-Based Harms

The digital exclusion of women and girls is not an accidental consequence of online abuse—it is often its intended effect. To ensure an effective Online Safety framework, Ofcom must adopt a regulatory approach that reflects the systemic, ideological, and escalating nature of online misogyny, particularly as it intersects with pathways to radicalisation and extremism.

Recommendations for Ofcom's Final Guidance

1. *Define Misogyny as Structural and Ideological*

   Misogyny should be recognised not simply as personal hostility, but as a mechanism of social enforcement—a system of norms, practices, and punishments directed at women who defy patriarchal expectations. It includes acts committed by both men and women and functions to reinforce gender hierarchy.

2. *Identify Male Supremacism as a Vector of Radicalisation*

   Platforms are incubators for male supremacist ideologies that glorify violence, promote rape culture, and recruit users into extremist groups. These dynamics must be treated as security threats with direct links to online and offline harm.

3. *Shift Risk Assessments from Individual to Systemic Harms*

Platforms should be required to assess and mitigate structural, cultural, and ideological risks, not just individual instances of abuse. Risk assessments must evaluate how platform architecture—such as algorithmic amplification, virality, and moderation failures—contribute to gender-based violence and exclusion.

4. *Mandate Transparency and Gender-Disaggregated Data Reporting*

   Platforms must publish publicly accessible, standardised data on abuse reports, content moderation, algorithmic decision-making, and survivor support. All data should be disaggregated by gender and other protected characteristics to enable intersectional analysis.

5. *Embed Safety-by-Design and Survivor-Centred Practices*

   Safety must be proactive, not reactive. Platforms should integrate trauma-informed, safety-by-design principles at every stage of development. This includes friction-reducing safeguards, red-teaming, usability testing with at-risk groups, and consultation with VAWG experts.

Framing the Issue within a Broader Rights Context

Online misogyny is not only a safety issue—it is a matter of human rights and democratic participation:

- *Freedom of Expression*: Abuse curtails women's right to speak and participate online.
- *Bodily and Psychological Safety*: Online abuse escalates to real-world harm including stalking, harassment, and sexual violence.
- *Democratic Integrity*: Gendered abuse suppresses women's voices in journalism, activism, politics, and public life.

Implementation Implications for Ofcom

To reflect the full scope of gender-based harms, Ofcom's guidance should:

- Broaden the taxonomy of harms beyond doxxing, deepfake abuse, gendered disinformation, and coordinated pile-ons.
- Recognise these as structural harms that shape the digital public sphere and entrench inequality.
- Apply a systems-level approach to platform accountability, emphasising that design, amplification, and inaction co-produce harm.

More effective accountability intervention: A Tax Proposal for Reduction of Harmful Online Content

I propose a "*pollution tax*" or Harmful Content tax (HCT) on online search and social media platforms.[1] This draws on the OECD's *Polluter Pays* principle, which is widely applied in environmental regulation. The "Polluter Pays" principle holds that those responsible for creating harm must bear the costs of mitigating its impact. The goal of a pollution tax is not primarily to raise large revenues, but to incentivize polluters to avoid tax by reducing pollution. Residual tax revenue has the benefit of paying for the regulatory costs, rather than placing the burden on the general tax payer.

How The Tax Would Work

- The tax would be levied on online platforms with a significant number of UK users.
- The revenue basis would be the advertising revenue generated, not profits.
- The rate of the tax would be determined by an estimate of the proportion of content exposures that contain harmful content. Harmful content exposure is how many users were exposed to harmful

[1] https://committees.parliament.uk/writtenevidence/132995/pdf/

content. Elements of the overall measure would include mis/dis-information and hate speech (e.g., Meta's 0.11% metric).[2]

How The Tax Rate Would be Calculated

The tax rate relies on estimating the amount of harmful content exposures as a proportion of all content exposures. This is technically feasible to estimate as Meta has already published estimates. Estimating the total pollution is a necessary element of any pollution tax.

We propose that the process of certifying the proportion of harmful content exposures would be carried out by certified third party providers. The analogy to use here is that of auditors preparing financial accounts for a company. Harmful content would be defined according to standards as would the auditing procedures. Content auditors would be certified against the standards.

The tax rate could be incremented in bands, or be proportional to the harmful content exposure rate (HECR). As an illustrative example of the latter, I assume a harm to tax rate coefficient of 50. If the HECR for a provider were 1 in 1000 (or 0.1%), the tax rate would be 5%. If the HECR were instead 1 in 100, the tax rate would be 50%. These rates of tax are on revenue. So if we assumed an average 10% tax rate (arising from an HECR of 1 in 500 and a coefficient of 50), the annual tax revenue (via approximating that platform providers of any size would be subject to the tax) would be of the order of £2.7Bn (using 2023 UK advertising spend with online search and social media platforms). These levels of tax are significant enough to induce the desired tax avoidance behaviour by platforms to reduce their tax by reducing their dissemination of harmful content.

Allocation of Tax Revenue

The tax aims to achieve two key goals:

- Create direct **economic incentives** for platforms to proactively mitigate the systemic harms their amplification mechanisms may generate and from which they may profit.
- Generate dedicated **funding** for crucial interventions like digital literacy programmes and supporting regulatory oversight functions, enhancing societal resilience.

The revenue generated from the HCT would be invested to achieve societal goods, e.g., by allocating it to:

- digital literacy programs to help users identify misinformation and fake news.
- independent regulatory bodies to fund oversight and research.
- law enforcement and organizations combating hate speech and misinformation.

Crucially, by focusing on the *economic activity* of amplifying harmful content for profit, rather than directly regulating speech content itself, the HCT might offer a pathway that addresses systemic harms while potentially navigating some of the direct free speech concerns tied to content moderation mandates.

[2] https://about.fb.com/news/2020/11/measuring-progress-combating-hate-speech/