ISD | Institute for Strategic Dialogue

Sponsored by:
Federal Foreign Office

Misogynistic Pathways to Radicalisation:

Recommended Measures for Platforms to Assess and Mitigate Online Gender-Based Violence

Sara Bundtzen

About the Digital Policy Lab

The Digital Policy Lab (DPL) is an inter-governmental working group focused on charting the policy path forward to prevent and counter the spread of disinformation, hate speech, extremist and terrorist content online. It comprises representatives of relevant ministries and regulatory bodies from liberal democracies. The DPL aims to foster inter-governmental exchange, provide policymakers and regulators with access to sector-leading expertise and research, and build an international community of practice around key challenges in the digital policy space. We thank the German Federal Foreign Office for their support for this project.

About this Paper

As part of the DPL, the Institute for Strategic Dialogue (ISD) organised working group meetings on the topic of online gender-based violence between May and June 2023. The working group comprised DPL members representing national ministries and regulators from Australia, Canada, Germany, New Zealand, the UK, and the US. Members of the Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership) and the Christchurch Call's multistakeholder Community joined the working group and contributed to this paper. Participants also included representatives from civil society, academia and industry.

While participants contributed to this publication, the views expressed in this paper do not necessarily reflect the views of all participants or any governments involved in this project.

ISD | Institute for Strategic Dialogue

Copyright © Institute for Strategic Dialogue (2023). The Institute for Strategic Dialogue (gGmbH) is registered with the Local Court of Berlin-Charlottenburg (HRB 207 328B). The Executive Director is Huberta von Voss. The address is: PO Box 80647, 10006 Berlin. All rights reserved.

www.isdgermany.org

About the Author

Sara Bundtzen is an Analyst at ISD, where she studies the spread of information manipulation by state and non-state political actors in multilingual online environments. As part of the Digital Policy Lab (DPL), Sara informs ISD's advisory work and analyses proposed pathways toward countering disinformation, influence campaigns, hate speech, and extremist content. Sara previously worked at the Federal Ministry of Defence and NATO HQ.

Contributing editor

Helena Schwertheim is a Senior Digital Policy and Research Manager at ISD Germany, where she leads the Digital Policy Lab (DPL). Helena is interested in government and multistakeholder responses to the digital threats posed to democracy by disinformation, hate speech and extremism. Previously, Helena managed digital policy and research projects at Democracy Reporting International, and she has experience in risk and political analysis in international organizations and think tanks, including at the UN World Food Programme, and International IDEA.

Editorial responsibility

Huberta von Voss, Executive Director, ISD Germany & Henry Tuck, Head of Digital Policy, ISD

Acknowledgements

We would like to thank all participants of the working group, including experts from governments, civil society, academia, and industry, for their contributions. We would like to give special thanks to the speakers as well as contributors to this paper for providing valuable insights and feedback: Elisabeth Brown (Christchurch Call Unit, Department of the Prime Minister and Cabinet (DPMC), New Zealand), Cailin Crockett (National Security Council / Gender Policy Council, The White House, US), Stella Ivory (Christchurch Call Unit, DPMC, New Zealand), Clara Martiny (ISD), David Reid (Christchurch Call Unit, DPMC, New Zealand), Sara Soleymani (Ministry for Women, New Zealand), Caroline Sinders (Convocation Research + Design), and Professor Lorna Woods (University of Essex, UK). We would like to thank ISD colleagues Milo Comerford, Cooper Gatewood, Melanie Smith, Dr. Tim Squirrell, and Henry Tuck who reviewed this paper.

Table of Contents

Executive Summary

This paper reviews online gender-based violence (OGBV) as existing within a continuum of (online and offline) violence, emphasising the connections with different extremist ideologies, including the dissemination of terrorist and violent extremist content (TVEC). It aims to prioritise a gender perspective in responding to TVEC so that social media platforms can better intervene in and mitigate misogynistic pathways to radicalisation that can begin (or be reinforced) online.

Given the scope of this review, focusing on platforms rather than broader forms of digital technologies, the paper uses the terminology OGBV in place of technology-facilitated gender-based violence (TFGBV).

The discussion asserts that the root causes of gender-based hate, misogyny, and other intersecting forms of identity-based hate and violence mirror a broader societal challenge that cannot be addressed or fixed by platforms alone. It thereby recognises that the mitigation of OGBV and online pathways to radicalisation requires a whole-of-society and whole-of-government approach. Whilst there are steps that governments and civil society can and should take, such as overseeing and enforcing emerging regulatory frameworks and voluntary commitments, this paper and its recommendations emphasise the role and actions of platforms.

Outlining the impact of OGBV at micro (individual) and macro (societal) levels, this paper considers how OGBV can be a vector for radicalisation, and is motivated by misogyny, which also pervades terrorist and violent extremist ideologies. The paper concentrates on the role platforms can play in exacerbating the risks of OGBV, evaluating platform policies, content moderation practices, user interface design and algorithmic recommender systems.

The discussion considers OGBV as any form of violence, including dehumanising language, directed against persons based on their gender identities or expressions, with intersecting protected characteristics such as (but not limited to) race, indigeneity, religion, sexual identity, class, or disability increasing the risks of experiencing OGBV. The paper recognises that women and LGBTQ+ people experience OGBV disproportionately. It situates OGBV as inherently linked to longstanding patriarchal gender norms, with misogyny functioning as an ideological link across a continuum of violence and as a vector across different extremist ideologies.

In this context, the paper asserts that researching and mitigating the risks of OGBV can enable earlier warning of and intervention in misogynistic pathways to different forms of violent extremism. Reiterating that any mitigation of risks must come in support of users' fundamental rights, including their right to privacy and freedom of expression, the paper proposes the following key recommendations.

Key recommendations

Enable API access to publicly available data for public interest research:

- The systematic collection of publicly available data via access to Application Programming Interfaces (APIs) can help complement digital ethnographic (and other) research methods, by filling in data gaps for the purpose of public interest research.

- Platforms should enable access to continuous, real-time or near real-time, and searchable APIs to allow vetted researchers to study the evolving tactics and forms of OGBV, as well as the links between online misogyny, radicalisation pathways and violent extremism. For example, such access could support longitudinal studies of in-group gender norms and behaviours over time, across different extremist ideologies and across platforms.

- Vetting processes of researchers should be inclusive to enable interdisciplinary research, involving a range of disciplines such as Computational Linguistics, Critical Terrorism Studies, and Critical Studies on Men and Masculinities, as well as recognising the value of comparative research across different ideologies, local contexts, and languages.

- While API access requires some form of vetting to prevent malicious or commercial uses, access should be free or at a nominal cost for researchers. Higher costs risk a de-facto inability to access data, or inequity among less well-resourced researchers.

Develop gender-disaggregated and standardised transparency reporting:

- Transparency reporting by platforms should enable external researchers to track and scrutinise the scope and scale of OGBV and the enforcement of community guidelines over time.
- Platforms should develop enforcement reports to include gender-disaggregated data, referring to statistical data in relation to community guideline violations. For example, platforms' processing of, and reporting on hate speech violations should include data on whether the violation was on the basis of gender and other intersecting protected characteristics, to allow intersectional analysis of the motivations driving OGBV.
- Platforms should work towards standardisation of transparency reporting through the development of a set of common metrics and content categories, whenever possible, to allow for comparison and tracking of policy violations across platforms. For example, standardised reporting should disclose the proportion of image-based versus text-based content that violated hate speech policies.
- A cross-platform effort to standardise reporting could coordinate and align with work by UN Women to develop a statistical framework for TFGBV and the UN Special Rapporteur on Freedom of Expression to develop a common definition for gendered disinformation.
- Platforms should consult and collaborate with GBV and feminist advocates, scholars, and victims-survivors with lived experience when developing the methodology of transparency reports or any internal research (e.g., when conducting user surveys). At a minimum, platforms should be transparent about the methodology of their reporting (and any changes thereof).

Apply a victim-survivor-centred Safety and Privacy by Design approach:

- Taking a victim-survivor-centred perspective, the development of user interfaces and tools should apply a gender and trauma-informed lens throughout all stages.
- Platforms should adopt proactive measures that support user agency with tools that protect their privacy and reduce exposure to OGBV; reactive measures that allow efficient user reporting (where possible, across platforms); and accountability measures that deter and sanction perpetrators appropriately.
- Data privacy and security should be embedded not only via accessible and transparent settings, but also in platforms' policies to moderate and mitigate the use of personal data for OGBV (e.g., to prevent doxing or sharing of intimate images without consent).
- Content moderation tools developed by industry such as Google's Perspective API should be continuously tested and scrutinised, taking into account a victim-survivor-centred perspective. Such efforts should be part of a cross-sector and multistakeholder dialogue.

Enhance cross-platform cooperation and information sharing of OGBV incidents (including actors and tactics):

- Platforms should develop and operate exchange channels between relevant teams, including content moderation teams, to proactively share information about OGBV incidents, including cross-platform harassment (such as relevant information about perpetrators using accounts across platforms). This is important for understanding the scope and scale of OGBV, but also to coordinate cross-plattform responses and mitigation actions, where appropriate.
- Platforms should share information about user reports, where appropriate, and develop interoperable reporting mechanisms, where possible, to support user agency.
- Existing cross-platform coordination such as the Christchurch Call Crisis Response Protocol and the Global Internet Forum to Counter Terrorism should review how OGBV is relevant to their mandates and adapt their scope and mechanisms appropriately.
- Cross-platform knowledge exchange should further build on and improve existing content moderation tools, including through regular assessments and reporting about the efficiency and impact of these tools.

Review content moderation policies, processes, and systems to acknowledge the continuum of violence and misogyny as a vector for violent extremism:

- Content moderation should account for the continuum of violence and recognise misogyny as a gateway and early warning sign of different extremist ideologies.
- Platforms should review and update hate speech and TVEC policies to recognise how misogynistic beliefs pervade ideological pathways to extremism. This includes considerations of how misogyny can be an ideology that encourages violent extremism and justifies violence towards women and LGBTQ+ people.
- Review processes should include the perspectives of women, the LGBTQ+ community, and victims-survivors. Such processes should also be inclusive of other protected characteristics, including race and religion, noting that GBV towards racialised communities often comes through the vehicle of racist and dehumanising language.
- Content moderation should account for veiled and coded misogynistic content, including contextual image-based content, as well as the multilingual, cross-cultural contexts of online spaces. For example, platforms could develop lexicons of words and phrases in cooperation with local organisations. Such efforts should be trauma-informed.
- The use of Artificial Intelligence (AI)-based systems for the purpose of detecting and moderating misogynistic content needs to be complemented by human oversight to allow for nuanced approaches that recognise the role of subtle and veiled misogyny, while also preventing false positives.

Apply intersectional feminist knowledge in risk assessments of AI-based systems:

- Platforms should incorporate gender analysis and feminist methodology when assessing the risks of algorithms and machine-learning (ML) models embedded in their services. This approach is useful for understanding how structural gender inequalities and patriarchal gender norms can be reproduced and amplified by AI-based systems.
- For example, platforms should review and update their recommendation guidelines (e.g., guidelines for content lowered in feeds) in alignment with a review of their community guidelines.
- Platforms should adopt victim-survivor-centric design processes from the ideation, conceptualisation, developing, testing, and scaling of new features or any changes to existing ones.
- Platforms should ensure that relevant teams (such as those designing, testing, and evaluating algorithms) are diverse and trained on how to conduct gender analysis to detect and mitigate biases and discriminatory patterns in their systems.

Strengthen and encourage multistakeholder dialogue and collaboration:

- As part of a broader good-faith effort, platforms should contribute to a trusted environment that supports exchange between stakeholders, including policymakers, government agencies, civil society, academia, product developers and software engineers.
- Multistakeholder exchange should actively seek intersectional perspectives, including those of victims-survivors of OGBV.
- Regular exchange could focus on testing and enhancing methods for assessing and mitigating OGBV. For example, stakeholders could discuss interoperable user reporting, content moderation tools, and algorithmic pathways.
- Multistakeholder collaborations should encourage the exchange of resources. Platforms and partners could consult the following resources for additional evidence and recommendations:
  - "Technology-facilitated gender-based violence: preliminary landscape analysis" by the Global Partnership for Action on Gender-Based Online Harassment and Abuse;

- "Technology Companies Must Make Platforms Safer for Women in Politics", "Interventions to End Online Violence Against Women in Politics" and "Landscape Tracker" by the National Democratic Institute (NDI);

- "The Chilling: A global study of online violence against women journalists" by the International Center for Journalists (ICFJ)/UNESCO;

- "Guidance on the Safe and Ethical Use of Technology to Address Gender-based Violence and Harmful Practices" by the UN Population Fund (UNFPA);

- "Measuring technology-facilitated gender-based violence. A discussion paper" by the UN Population Fund (UNFPA);

- "Report on freedom of expression and the gender dimensions of disinformation" by the UN Special Rapporteur on the promotion and protection of freedom of opinion and expression;

- "Technology-facilitated violence against women: Taking stock of evidence and data collection" by UN Women;

- "Violence Against Women and Girls (VAWG) Code of Practice" by The End Violence Against Women Coalition, Glitch, Refuge, Carnegie UK, NSPCC, 5Rights, Professor Clare McGlynn and Professor Lorna Woods.

Glossary

Application Programming Interfaces (APIs) are software intermediaries that allow two applications to communicate with each other. APIs have a huge range of uses, but in the context of this Explainer, they allow researchers to access certain types of data from some online platforms via requests. As an intermediary, APIs also provide an additional layer of security by not allowing direct access to data, alongside logging, managing and controlling the volume and frequency of requests.

Extremism is the advocacy of a system of belief that claims the superiority and dominance of one identity-based 'in-group' over all 'out-groups'. It propagates a dehumanising, 'othering' mind-set incompatible with pluralism and the universal application of human rights. According to ISD's definition, extremism can manifest through violence and the targeting of hate towards groups on the basis of their identity, as well as more gradualist supremacist social or political projects that undermine human rights, democratic institutions and civic culture. It is important to place OGBV on the spectrum of extremism as misogynistic content is used in radicalisation processes, and can incite and translate to offline violence. **Violent extremism** is understood in this context as a specific violent manifestation of the wider phenomenon of extremism.

Gender is an individual's internal sense of being a woman, a man, neither of these, both or somewhere along a spectrum.[1] It describes socially constructed roles for women and men, and is an acquired identity that is learned, changes over time and varies widely within and between cultures. **Gender norms** or **gender stereotypes** are "generalised views or preconceptions about attributes or characteristics, or the roles that are or ought to be possessed by, or performed by, women and men."[2] They are often framed in a binary that overlooks the lived experience and richness of gender-diverse people, while also being trans exclusionary. In contrast, **sex** is assigned at birth based on the physical appearance associated with being female or male.

Gender-based violence (GBV) refers to "violence directed against a person because of that person's gender or violence that affects persons of a particular gender disproportionately."[3] Women and LGBTQ+ community, including transgender and gender-diverse people, experience disproportionate rates of GBV.

Male supremacy is a "hateful ideology rooted in the belief of the innate superiority of cisgender men and their right to subjugate women [and trans and gender-diverse people]."[4] It is linked to **hegemonic masculinity**, which structures patriarchy and describes the "legitimation of unequal gender relations."[5]

The **Manosphere** is an umbrella term that refers to several interconnected misogynistic communities online. It encompasses multiple types and severities of misogyny with varying expressions of violence – from broader male supremacist discourse to Pick Up Artists, Men's Rights Activists (MRAs), Men Going Their Own Way (MGTOW), and involuntary celibates (incels).[6]

Misogyny operates to uphold a patriarchal social order, policing gender norms to ensure that women and marginalised gender identities conform.[7] It works to justify violence if these norms are deviated from.[8] Misogyny includes what might be considered a type of deeply held prejudice towards women and marginalised gender identities and intersects closely with racism, antisemitism, Islamophobia, ableism, and anti-LGBTQ+ hate. Misogyny thereby operates alongside other intersecting forms of discrimination, including misogyny targeted at transwomen (transmisogyny) and the specific form of hatred Black women face (misogynoir[9]). It is often hidden within different forms of violent extremist ideologies. It is also a motivating ideology in itself, separate from other types of extremist ideologies.[10]

Online gender-based violence (OGBV) can be described as a subset of **technology-facilitated gender-based violence (TFGBV)**, which refers to any "act that is committed, assisted, aggravated, or amplified by the use of information communication technologies or other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political, or economic harm, or other infringements of rights and freedoms."[11] The phenomenon is also referred to as **technology-facilitated violence against women (TFVAW)**, noting that VAW can be substituted with GBV, whilst maintaining the common definition describing the phenomenon.

Radicalisation is a term used in this context to describe the process by which an individual adopts an extremist ideology (defined above), which may (or may not) enable acts of violent extremism or terrorism. In the literature on terrorism and violent extremism specifically, a frequent distinction is made between cognitive radicalisation (adopting extremist beliefs) and behavioural radicalisation (the process leading up to violent behaviour).

Introduction

Gender-based violence (GBV) in public and private life is a global challenge that has been increasingly connected to and amplified by the online spaces of social media platforms, messaging services and other communications technologies. It reflects the manifestation and amplification of unequal power relationships that stem from patriarchal gender norms,[12] which can be directed at all genders, but most often towards women and LGBTQ+ people. It also intersects with other forms of identity-based violence such as (but not limited to) racism, Islamophobia, and antisemitism.

While social media platforms can help empower feminist movements, for example, by bringing greater visibility and attention to women's and LGBTQ+ communities' rights issues, the current online environment can enable and reinforce misogynistic and anti-LGBTQ+ content. Further, OGBV disproportionately affects women in public life, including activists and human rights defenders,[13] politicians,[14] and journalists[15] causing a 'chilling effect' on equal civic and political participation – with gendered and sexualised mis- and disinformation also being used as deliberate tactics by (both non-state and state) anti-democratic actors.[16]

The level of response by platforms to address misogyny on their services varies, and some have taken commendable actions. However, so far, no platform has identified and taken sufficient steps to effectively address the individual and societal risks emanating from OGBV.

While online manifestations of GBV have distinct features, they belong to a "continuum of multiple, recurring and interrelated forms of GBV."[17] OGBV enforces and amplifies the patriarchal order with tools from across a tactical spectrum, ranging from legal but harmful behaviour to terrorism and violent extremism.[18] GBV manifests online, while the reproduction and amplification of misogyny online can lead to offline violence – ranging from intimate partner violence, physical attacks against female journalists to mass violence.[19]

In this context, is important to recognise that OGBV occurs within an ecosystem characterised by a gender digital divide that is rooted in structural gender inequalities, in which those who design – and, in some countries or regions, access and use – communication technologies are disproportionately male.[20] Recognising the intricate dynamics of online and offline GBV, this paper elucidates the connection of OGBV with extremist ideologies and violent extremism, with a focus on evaluating and addressing the role and actions of platforms.

A range of multilateral fora including UN Women, the World Health Organisation (WHO), and the UN Population Fund (UNFPA), as well as multistakeholder initiatives such as the Global Partnership recognise the need for action to address the role of platforms in enabling and exacerbating OGBV. The Christchurch Call has recognised the need to deepen and explain the evidence base on the links between misogyny and TVEC.

Drawing on discussions with government, industry and civil society stakeholders, this paper reviews the trends and multi-level impacts of OGBV, emphasising the multifaceted relationship with violent extremism. Based on this review, the paper proposes risk assessment and mitigation measures for platforms to respond to misogynistic content and behaviour on their services.

Trends in online gender-based violence

OGBV has tangible and measurable offline impacts, and offline harms can be extended and amplified online. In many cases, the victim-survivor knows the perpetrator, who is often a current or former partner, relative, co-worker, or friend.[21] Recognising the commonality of gendered power relations as elements of both intimate partner violence and extremist ideologies, the following section examines how misogyny, including gendered and sexualised motives and attitudes, overlaps with or becomes a vector for violent extremism. It acknowledges the need to consider the intersectionality of OGBV with other forms of violence such as racism and Islamophobia, as well as the relationship of online/offline manifestations of violent extremism.

Actors: OGBV as a vector for radicalisation and violent extremism

In recent years, there has been an increasing focus on gender dynamics in the context of research studying online radicalisation and the dissemination of TVEC across different extremist ideologies.[22] Some scholars have noted the need for further recognition of misogyny as an ideological vector for radicalisation in Preventing and Countering Violent Extremism (P/CVE) programming and policy.[23]

A growing body of research focuses on the relationship between far-right extremism, misogynistic ideology and the Manosphere; the latter being a loose network of misogynistic online communities that seek to enforce male supremacy and patriarchal gender norms.[24] For example, the Southern Poverty Law Center (SPLC) explains that male supremacists "are fixated on rigid gender roles and vilify any deviation from their strict gender dichotomy," describing male supremacy as a "powerful undercurrent for white supremacy, and its tenets undergird much of the contemporary far right."[25] Misogynistic violence has manifested in physical attacks on women, with misogynistic motivations also intersecting with racist and xenophobic sentiments. For example, the perpetrator of the spa shootings in Atlanta, Georgia in 2021, killing six women of Asian descent, displayed "gross misrepresentations of hypersexualized Asian women."[26]

However, the relationship between misogynistic groups such as 'involuntarily celibates' (incels) – a subset of the Manosphere who blame women and society for their lack of romantic success[27] – and other supremacist ideologies is complex and multifaceted. For example, incels possess a unique perspective on race and ethnicity that differs from far-right groups, involving a perceived racial hierarchy in the dating sphere favouring white men, which they attribute to female choices in selecting sexual partners[28] (rather than actively endorsing it). Far-right groups meanwhile drive a more racially supremacist vision, looking to enforce racialised sexual boundaries to maintain in-group homogeneity. In recognition of these nuances, there has been cross-pollination as both, incels and far-right groups, share a misogynistic ideology and antifeminist sentiments. ISD research notes that some incels explicitly identify with racially or ethnically motivated violent extremism (REMVE), for example, by labelling themselves 'stormcels' in reference to Stormfront, a notorious white supremacist website.[29]

In this context, P/CVE policy and programming has lacked a focus on the potentially violent extremist outcomes of misogynistic ideology. Notably, recent scholarly debates have concentrated on whether examples of incel associated violence should be understood as constituting terrorism.[30] While incels are certainly political in nature with a core ethos geared towards subjugating and repressing a group of people, there is no consensus on whether violence by this group should be considered primarily ideological, or alternatively nihilistic. In this context new legal precedents are being set, with the Ontario Superior Court of Justice recently determining an incel-motivated murder amounted to terrorist activity.[31] In such instances, notions of 'lone-wolf' actors can be a misnomer. While perpetrators of GBV, both online and offline, might not affiliate to a particular group, this may be due to the nature of misogynistic ideology as diffuse, networked and pervasive, a phenomenon related to the wider challenge of 'post-organisational' extremism.[32]

Misogyny – like antisemitism – often serves as a unifying core feature of different extremist ideologies. In the context of promoting hetero-normative gender norms and identities, the connections between violent extremism and misogyny showcase parallels between militant masculinity in different ideologies, notably in far-right and Islamist extremist groups. Researchers note that both "equate manliness with the readiness

to defend and it is not uncommon for overt or diffuse misogyny to serve as a motivating force for turning to the respective ideology."[33] Examining violent Islamist extremist actors, researchers observed that ISIS had practiced "a militarised, masculinised, religious and genocidal nationalism within their 'Islamic State' when subjecting Yazidi women and girls and other minorities to GBV."[34] Both Islamist and far-right extremism "impose patriarchal gendered roles, binaries, hierarchies, and norms,"[35] reiterating that male supremacist and misogynistic belief systems are present across a diverse ideological spectrum. Common to all is their misuse and exploitation of mainstream and 'alternative' platforms and messaging services, cutting across geographical locations and languages.

In addition, researchers highlight how antisemitism intersects with misogynistic beliefs. Evelyn Torton Beck explains that the 'Jewish Princess' stereotype "remodels the traditional antisemitic tropes onto a female form: she is materialistic, money-grabbing, manipulative, shallow, crafty and ostentatious."[36] Blyth Crawford notes that the neofascist militant accelerationist movement sees Jewish people "as influencing sexual politics in ways that are regarded as being 'anti-family' and therefore constitute a threat to the white race."[37]

Part of the 'white genocide' conspiracy theory, there are connections between antisemitic tropes and anti-LGBTQ+ fearmongering, involving gender-based hate speech. The Anti-Defamation League (ADL) asserts that the "alleged targeted promotion of LGBTQ+ identities and relationships is seen as a key element of Jews' attempts to reduce reproduction rates among straight, cisgender white people."[38] Anti-LGBTQ+ tropes labelling the community as "pedophiles" or "groomers"[39] often converge with the antisemitic canard that Jews prey upon non-Jews, especially non-Jewish children.[40] This observed confluence of antisemitism, misogyny and anti-LGBTQ+ hate is perpetuated by a wide range of extremists with different ideological backgrounds.

Structural gender inequality and gender norms can also lead to internalised misogyny. An extreme example is the emergence of Tradwives as an influential online community, showcasing the reinforcing elements of far-right ideology, Christian Nationalism, white supremacy, and patriarchal gender norms. Tradwives embrace a highly hetero-normative rendition of the 'wife and mother' role, in opposition to feminism, reproductive rights, LGBTQIA+ rights, and gender equality. Researchers highlight that Tradwives "use their presence on social media to offer a powerful female in-group association" and "successfully infiltrated mainstream social media with their anti-globalist, anti-modern approach to life."[41] While women's role in violent extremist ideologies and communities varies, research finds that women-only forums have also "served as gendered sites of ideological contestation," where women are asserting "agency in their everyday practices despite otherwise constraining gendered ideological constructs."[42]

While extremist communities often use veiled and coded language to conceal and convey in-group culture,[43] online misogyny can become widespread and popular. ISD finds that it is most impactful and prolific when different ideological groups participate in the spread of misogyny. Anti-drag and anti-LGBTQ+ activities, for example, are not limited to fringe groups, but have become a unifying concern for the far-right as well as localised activists, including certain parents' rights groups, anti-vaccine or anti-lockdown groups, and Christian nationalists.[44] Additionally, OGBV and misogyny has been highly associated with violent conspiracy movements. For example, ISD research on the online activities of QAnon supporters has shown how targeted hate, including violent misogynistic, racist, and anti-LGBTQ+ rhetoric, has become a particular concern for prominent women, who often found themselves on the receiving end of coordinated harassment.[45] Finally, ISD research shows that a small group of actors can have considerable influence over the propagation of misogynistic content, including for example prominent influencers like Andrew Tate.[46]

Misogynistic content can thereby serve as an ideological link across different extremist groups, with increased exposure to online misogyny risking a normalisation among users, especially among male users who use online spaces to socialise, network, and connect with others.

Tactics: OGBV as a continuum of violence

As stated, OGBV occurs within a continuum of violence,[47] which recognises the complex and interlinked experiences of different forms of violence.[48]

This section outlines prevalent forms of OGBV, reviewing a 2020 survey on 'Measuring the prevalence of online violence against women' conducted by the Economist Intelligence Unit (EIU)[49] as well as other relevant research, including by PEN America,[50] the Global Partnership,[51] and UNESCO/International Center for Journalists (ICFJ).[52] It is not intended to be exhaustive, but to indicate the myriad forms of OGBV. Importantly, these forms and behaviours are frequently observed in combination and across multiple platforms.

- **Online gendered or sexualised mis- and disinformation** refers to "a subset of online gendered abuse that uses false or misleading gender and sex-based narratives against women [and trans and gender-diverse people], often with some degree of coordination, aimed at deterring women [and trans and gender-diverse people] from participating in the public sphere."[53] It may involve defamatory comments that intend to harm a person's reputation. A combination of false information with the publication of factual, decontextualised and misrepresentative information is often the most harmful. Gendered and sexualised mis- and disinformation often uses coded and veiled language as well as iterative, context-based visual and textual memes.

- **Online harassment** encompasses a wide range of unwanted or negative contact that is used to create an intimidating, annoying, frightening, or even hostile environment.[54] It can involve long-lasting coordinated narrative framing, sharing of target lists, and brigading across platforms.[55] It may also be in the form of a single comment or one-off incident. It is often gendered or sexualised in nature.

- **Online gender-based hate speech** attacks or humiliates persons based on their gender identities and expressions, with intersecting identity factors such as (but not limited to) sexual identity, ethnicity, race, religion, or disability increasing risks of becoming a target of hate speech.[56] It can range from dehumanising and derogatory language to threats and incitements of violence.[57]

- **Online impersonation** refers to wrongfully obtaining and using another person's personal data in some way that involves fraud or deception. Gendered examples include creating fake accounts to groom and recruit girls and women into sex trafficking.[58]

- **Stalking and monitoring** involve the misuse of technology, such as installing commercial stalkerware on a device. Stalking and monitoring is often repeated, and can be an extension of intimate partner violence.[59]

- **Astroturfing** refers to the deceptive practice of dissemination or amplification of content that appears to arise organically at the grassroots level, but is actually coordinated by an individual, interest group, political party, or organisation.[60] Astroturfing may be part of **networked harassment**, which involves tactics such as **trolling** (purposely upsetting or disrupting online events, debates or hashtags)[61] and **coordinated flagging** (falsely reporting users to get them de-platformed).[62]

- **Image-based sexual abuse** involves the creation, distribution, sharing or threat of sharing intimate images or videos of a person without their consent.[63] It includes a diversity of behaviours such as **sexual extortion** (when a person has, or claims to have, a sexual image of another person and uses it to coerce them into doing something they do not want to do);[64] **documentation or broadcasting of sexual violence** posted on social media, texted among peers, sold or traded, resulting in an additional form of sexual violence against the victim-survivor;[65] and the **use of generative AI to construct deepfakes**, including artificial images or videos that resemble actual photographs or videotapes.[66]

- **Doxing** involves retrieving and publishing of personal or identifying information (e.g., addresses, phone numbers, emails, partners' or children's names) without permission – often with a malign intent to show up at the workplace or home, or to make negative or unwanted contact.[67]

- **Threats of offline violence** such as rape and death threats, or incitement to physical violence. Women journalists,[68] academics,[69] politicians[70] and human rights defenders[71] often face violent threats, which are gendered and sexualised, particularly if they are speaking or writing about equality issues or male-dominated topics. A global study conducted by IFCJ mapped the vicious circular trajectory of online violence, highlighting that "digital attacks can fuel offline violence, while offline abuse by prominent figures can trigger online pile-ons."[72]

These forms of OGBV occur across platforms, often simultaneously and in a coordinated manner. Astroturfing and networked harassment tactics misuse platforms to facilitate wider reach of misogynistic content as well as the networking and in-group building of perpetrators who might otherwise be isolated from one another. For example, incel forums are spread across Reddit and 4chan as well as gaming forums like Discord or dedicated websites, which reinforce in-group, community and belonging.[73] While this type of cross-platform misogynistic behaviour and networking creates additional risks for users, it also reiterates the challenge of understanding how vulnerable individuals become radicalised and how a healthy online environment can help prevent this from happening, including tailored interventions that address individual grievances.

Moreover, coordinated harassment campaigns often take advantage of online conversations surrounding trending topics, which may involve the use of abusive hashtags, to spread misogynistic content.[74] This can also manifest in the coordinated harassment of an individual across multiple platforms. In turn, coordinated online harassment raises challenges for the tracking and reporting of OGBV, often putting the onus on victims-survivors, and illustrating the need for victim-survivor-centric coordination and collaboration between platforms.

Finally, the inherently global reach of many platforms expands the online misogynistic influences in radicalisation pathways, contributing to a perpetuation of online cultures of extremist beliefs. For example, studies have shown that occasional encounters with extremist content are experienced by 40% to 50% of younger individuals.[75] This creates constant opportunities for the initiation of radicalisation processes within large populations.

Harms: impact of OGBV on individuals and society

This section outlines the impact of OGBV at both the micro (individual) and macro (societal) level, recognising the range of harms, including the risks to private as well as public safety.

- **Psychological harms:** Research shows that OGBV can leave victims-survivors with serious psychological harms, mental or emotional stress, as well as symptoms of post-traumatic stress disorder, particularly when the abuse is frequent. Cumulative effects of offline and online violence can also lead to self-harm, depression, and suicide. [76]

- **Threats to reproductive health:** Misinformation about abortion and reproductive rights can cause gendered harm as it undermines access to correct information about health care and promotes unsafe alternatives or unproven medication.[77]

- **Privacy invasions:** Once personal information is released online (for example via doxing), it can be difficult, if not impossible, to retrieve or remove. This also create risks of future invasions given that the information remains permanently on the Internet or stored on another person's device. This negatively impacts the right to privacy of targeted persons.

- **Economic and material harm:** The term 'economic vandalism' highlights the economic costs caused by OGBV, for example, due to missed work opportunities, decreased productivity, and retreating from the Internet.[78]

- **Exacerbating structural gender inequality:** OGBV normalises misogyny and promotes a culture of patriarchal violence, involving rape culture, victim-survivor blaming and trivialising sexual assault.[79] The normalisation of OGBV reinforces a 'silencing' of women and LGBTQ+ people, whereby the victim-survivor is discouraged from participating in public life. OGBV thereby exacerbates gender inequality that limits women and LGBTQ+ community from exercising their freedoms and human rights. There is also an intergenerational impact as OGBV deters and impedes young women and girls and LGBTQ+ people entering professions such as politics and journalism, due to fear of similar abuse, which, in turn, increases the gender digital divide.[80]

- **Threats to private and public safety:** Reiterating that OGBV occurs within a continuum, misogynistic behaviour that starts in the online space may lead to the perpetration of offline violence – both in private and public spheres.[81] For example, a 2023 US Secret Service report details the public security threat posed by individuals who perpetrate acts of targeted violence, with attackers engaging in domestic violence, misogynistic behaviours, or both prior to an attack. It notes that men who have committed misogynistic violence (typically mass shooting and stabbings) have histories of concerning and threatening online communications, as well as other risk factors (such as a history of being bullied, financial instability, and interpersonal difficulties).[82]

- **Threats to democracy:** At a societal and global level, anti-democratic forces – both foreign state and non-state malign actors – exploit online spaces to attack women and LGBTQ+ people in public life.[83] A report by #ShePersisted notes that gendered disinformation can serve as an early-warning system for "both backsliding on women's rights and the erosion of democratic principles and institutions."[84] A global study conducted by ICFJ notes an "alarming trend" of the role played by political actors, including politicians, government officials, political party representatives, party members, political operatives, and extremists on the political fringe, as "instigators and primary perpetrators of online violence against women journalists."[85] A global report on gendered disinformation by the U.S., Canada, the European External Action Service, Germany, Slovakia, and the UK further emphasises that foreign state actors like Russia and the People's Republic of China (PRC) strategically target women and people with intersecting identities to dissuade individuals and identity-based groups from exercising their rights. The report further asserts that identity-based disinformation undermines the "ability to access impartial, fact-based information, and it negatively impacts the make-up of democratic representation."[86]

Assessment of platform policies and enforcement

A key factor in responding to OGBV is the development and effective enforcement of comprehensive community guidelines or standards, which outline what is and what is not allowed on a platform. These are generally contracts of adhesion, presented to users on a take-it-or-leave-it basis, and include a set of policies that are frequently updated by the platforms.

While most platforms[87] generally account for some forms of OGBV in their hate speech or harassment and abuse policies (including the protected characteristics of users), this section outlines some of the gaps in both policy and enforcement. In terms of the former, none of the platforms explicitly address gendered or sexualised mis- and disinformation. Yet, such content often comes in the form of coded and veiled language, context-based visual and textual memes, or use tactics of intentionally obscuring certain words.[88] Furthermore, the Oversight Board, which reviews content decisions made by Meta, recently overturned Meta's decision to keep online a Facebook post that mocks a victim-survivor of GBV. Specifically, the Board found that the post violated Meta's Bullying and Harassment policy as it mocked the serious physical injury of the woman depicted. The Board explained, however, that "this post would not have violated Meta's rules if the woman depicted was not identifiable, or if the same caption had accompanied a picture of a fictional character," indicating a gap in policy that seems to allow content that normalises GBV.[89]

Beyond gaps in policy formulation, ISD research identified patchy enforcement of existing policies. On X (formerly Twitter), which prohibits "targeting others with repeated slurs, tropes or other content that intends to degrade or reinforce negative or harmful stereotypes about a protected category,"[90] ISD research conducted in the US context found multiple instances where this type of content was not moderated, including tweets containing sexist tropes against the actress Amber Heard as well as general attacks on women's appearances.[91] The same research also identified openly derogatory terms such as "whore", "cunt" or "bitch" in the comment section of YouTube, which would often occur not only under videos that seemed to invite hateful comments but also under inconspicuous videos (e.g., such comments were found beneath both "Andrew Tate Destroys Modern Women" and "House Speaker Nancy Pelosi holds her final weekly press conference").[92] On TikTok, ISD found that misogynistic content is still openly posted and promoted by users, including videos by accounts promoting Men Going Their Own Way (MGTOW), a sub-movement of the Manosphere, that among other things, belittles and dehumanises single mothers and their role in society.[93]

The platform with the most lax (or lacking) policies and enforcement (i.e., where ISD found the most misogynistic content) is Telegram. Notably, 'private' channels (which require an invite by the owner or an invite link to join, but in practice are often easily joined) are not covered by the terms of service. As of the time of writing, the service has no policies addressing hate speech, nor does it prohibit doxing, despite having been criticised for hosting "an epidemic of politically motivated doxing, allowing dangerous content to proliferate, leading to intimidation, violence, and deaths."[94] Researchers noted that while Telegram was designed as a messenger service, it has become a hybrid between a messenger service and a social media platform as messages in public channels can reach hundreds of thousands (or even million) views, with some channels enabling commenting on posts or reacting to them via emojis.[95] Moreover, ISD found that sharing links to other audio-visual platforms, such as video and livestreaming websites, is particularly popular among right-wing extremists and conspiracy theorists on Telegram.[96] This reiterates the challenge of cross-platform dissemination of content, coordination and activities, including the relevance of accounting for the risks posed by smaller and 'alternative' platforms.

Finally, a well-observed problem is the lack of sufficient moderation resources for content published diverse local contexts, including in non-Western regions or non-English languages. For example, a report by the Slovak Council for Media Services and Reset reviewed the role of platforms in the case of a shooting outside an LGBTQ+ bar in Bratislava in 2022, finding that Facebook had sent reported content to a hired third-party fact-checker to perform the review. However, there was only one Facebook-contracted fact-checker for all of Slovakia, showcasing how limited resources pose obstacles for the rapid and efficient review of content violating the community guidelines.[97]

Platform design and systems: Risks of reproducing and amplifying OGBV

Studying the impact of platform design and systems on exacerbating the risks of OGBV and the dissemination of misogynistic content remains a challenge given that users experience highly personalised online interfaces and spaces. This section reviews how platform design and systems risk reproducing and amplifying OGBV, assessing user interface design and artificial intelligence (AI)-based systems, including the use of algorithms and machine learning (ML) models for personalised feeds and (search) recommendations.

User interface design: A relatively well-studied phenomenon showcasing how design can be harmful to users are deceptive patterns, also known as "dark patterns". There are many forms of deceptive design patterns,[98] but essentially, they are "choices that can [unintentionally or intentionally] influence or trick users into making unintended decisions."[99]

In 2021, interdisciplinary research conducted by Caroline Sinders, Vandinika Shukla and Elyse Voegeli surveyed journalists' interactions and relationships with platforms' user interfaces, given their status as a user group that faces a range of harassment and harm online, including gendered or racial slurs, doxing, and rape threats.[100] Their research emphasises that "technology is a planned space, and users can only conduct specific actions that are designed and allowed by the software, application or platform they are using."[101] The design of online spaces is thereby closely interlinked with the experience of potential harms on platforms. Deceptive design choices can negatively impact users' privacy and safety, for example:

- Settings that default to the least privacy friendly option;
- Rewards and restrictions if users decline or opt out of settings, such as loss of functionalities;
- Forced action to complete the settings review at a time determined by the platform, pressuring users without a clear option to postpone the process;
- An illusion of control as the platform provides users with granular choices that ultimately discourage them from changing or taking control of their settings.

Deceptive design highlights the importance of applying a Safety by Design approach, which encourages platforms to build safety into the design, development, and deployment of their features, rather than retrofitting safety solutions after harms have occurred.[102]

The survey also indicates a lack of victim-survivor-centred reporting mechanisms and communication of community guidelines. Surveyed journalists experienced harassments in peaks, with patterns of harassment instead of stand-alone instances. However, they were only able to report individual instances rather than multiple instances in bulk. Journalists also expressed "frustration and confusion over how platforms responded to harassing content." It is important to consider that limited user agency in the face of this type of harassment may replicate a loss of power, control, and rights, which is an experience shared by many victims-survivors of OGBV. The coordination of harassment campaigns across platforms and the lack of interoperable reporting mechanisms further weaken user agency and risk mitigation.

AI-based systems: Personalised feeds and other AI-based services such as search recommendations shape the user experience, creating specific risks of OGBV. Algorithmic systems make automated decisions that score and rank content and suggestions for who to connect with (or what pages or groups to follow) based on signals, including users' historical behaviour (such as viewing history) and predictions derived from past behaviour of similar users (using techniques such as collaborative filtering).[103] ML models use predicted probability of engagement (the probability of users liking, sharing, viewing, etc. content) to optimise the order in which content is ranked and displayed on user feeds. In short, the goal of engagement-based ranking is to maximise whatever engagement goal (metric) a company has set, often at the level of individual user indicators (for example, the time users spend on a platform).

While studying the impact of algorithmic feeds on discourses and user behaviour remains challenging, there is evidence pointing toward an engagement problem, which describes the tendency to engage more with content that is low-quality (such as clickbait

headlines) or nears a "cut-off point" of what is allowed under the community guidelines (borderline content).[104] Therefore, engagement-ranked feeds often risk creating a "gravitation" towards borderline content, which also increases the risks of recommending misogynistic content.[105]

For example, ISD research suggests that platforms give greater visibility to abusive hashtags over non-abusive hashtags. On Instagram, the transphobic hashtag #rachellevineisaman (25 posts) was ranked third among recommendations, ahead of non-abusive hashtags featured in more posts (e.g., #rachellevinephotography with 183 posts).[106] Moreover, research finds that Tradwives are able to "adapt their content" to exploit algorithmic feeds, using self-branding strategies, presenting the "#tradlife" and sharing "homemaking, cleaning, and beauty content" to engage audiences, while promoting antifeminist and anti-LGBTQ+ belief systems. They further capitalise on engagement-based ranking by commodifying far-right ideology through advertising, brand collaboration, or promotions.[107]

Studying AI-based systems in general, scholars suggests that algorithms reflect and exacerbate gender norms already present in society. Patterns of gender bias and discrimination have been detected in algorithms used for hiring decisions,[108] criminal sentencing,[109] and health-care allocation,[110] among others. Researchers from New York University (NYU) demonstrated that societal levels of inequality are evident in search algorithms, noting a "cycle of bias propagation between society and AI."[111] This reiterates the need to address the challenges of AI for society, including how risks reproducing and amplifying OGBV, in a multistakeholder effort that involves interdisciplinary research and development of offline and online responses.

As outlined, platforms employ algorithms and ML models to make predictions about user engagement, using large amounts of user data. However, automated inferences risk biased outputs, including gender stereotyping, given that datasets often contain racial and gender biases.[112] For example, researchers tested the accuracy of X's (formerly Twitter) inferences of users' gender identities, finding that the LGBTQ+ community and straight women were more often misgendered than straight men. Researchers emphasise that misgendering users, "beyond echoing deeply rooted stereotypes, can lead to privacy and discrimination issues."[113]

Rebekah Tromble, Director of George Washington University's Institute for Data, Democracy and Politics, describes the problem this way: "how we consume social media content is an inherent human construct. And if there are problems with how this consumption happens, it's down to concerted decisions from social media executives and engineers — and not some natural phenomenon that is out of anyone's hands."[114] How platforms design and evaluate their algorithms directly impacts user experiences, especially as they risk amplifying the dissemination of misogynistic content.

Response measures: Risk assessment and mitigation of OGBV

Based on the review of trends in OGBV and platform policies, design and systems, the following sections evaluate and propose how platforms can work toward better assessing and mitigating OGBV and its connections with violent extremism. Recognising the need for multistakeholder approaches and solutions, this section considers proposed measures at the multilateral and governmental level, as well as by industry, civil society, and academia.

Enable API access to platform data and develop standardised transparency reporting

The monitoring, measurement, and transparent reporting of OGBV by platforms are prerequisites to understand and explain the nature, scale, and scope of the phenomenon. Additionally, API access to publicly accessible data[115] should support public interest research[116] and enable evidence-based decision-making.[117]

The Global Partnership together with UN Women, the WHO, UNFPA and UNICEF initiated efforts toward enabling the production of accurate, reliable and comparable data and knowledge around OGBV.[118] In 2023, the UN Women-WHO Joint programme on Violence against Women (VAW) data published a paper on 'Taking stock of evidence and data collection', scoping methodologies and recommendations on the approaches to collecting data on Technology-Facilitated Violence against Women (TFVAW).[119] The paper highlights existing methodologies[120] as well as methodological, ethical and socio-political challenges. These include the lack of "overall problematisation and awareness" around TFVAW due to a lack of data and dissemination of research findings, and a bias of data towards the Global North, neglecting the differentiated impacts across diverse and different contexts. The paper highlights the need for a shared operational definition and methodology for monitoring, measuring and analysing TFVAW. It further notes the importance of incorporating social media data and the need to consider a "diversity of methodologies" to allow for different data sources.

In this context, the collection and analysis of platform data should address data gaps for the purpose of evidencing the tactics and forms of OGBV as well as the connections between misogyny and different extremist ideologies. For example, API access could allow cross-ideological analysis of different violent extremist and terrorist actors, including a comparative analysis of the respective gender dimensions.[121] Vetted researchers from different disciplines such as Computational Linguistics, Critical Terrorism Studies, or Critical Studies on Men and Masculinities should have meaningful API access to systematically collect and analyse data. Regulatory frameworks already address the need for such data access. Notably, Article 40 of the European Union (EU)'s Digital Services Act (DSA) requires that access to data "publicly accessible in their online interface" should be made available, where possible, in real-time to researchers, including those affiliated to not-for-profit bodies, organisations and associations. In parallel, company signatories of the 2022 Strengthened Code of Practice on Disinformation committed to voluntary standards that will serve as co-regulatory measures for the DSA. The Code includes the commitment to "continuous, real-time or near real-time, searchable stable access to non-personal data and anonymised, aggregated, or manifestly-made public data for research purposes on Disinformation through automated means such as APIs."[122] Available platform data should be compared and triangulated with data from other sources such as administrative data, statistics, or surveys to ensure a comprehensive mapping of the phenomena.

In addition, platforms should develop standardised transparency reporting to include gender-disaggregated data to allow external researchers to scrutinise and track the enforcement of policies, especially considering violations of hate speech and TVEC policies. For example, enforcement reports should include aggregated data on the prevalence of and user engagement with content (including but limited to posts, comments, and profiles) detected as gender-based hate speech, the proportion of image-based content that violated these policies, as well as data on how user reporting was addressed (e.g., what specific actions were taken).[123] Enforcement reports should also account for hate speech directed towards other protected groups to measure intersecting identity-based hate and support intersectional analysis of the motivations driving OGBV.

Platforms should work with GBV and feminist advocates, scholars, and victims-survivors of OGBV when developing methodologies of transparency reports

(including content categories and metrics), or when conducting internal research (such as user surveys about experiences of OGBV). Considering the lack of a universally agreed definition of OGBV and the need for more consistency of transparency reporting (and thereby comparability of platform actions), a cross-sector effort could also contribute to and participate in the ongoing work by UN Women[124] and to develop terminology and a statistical framework for TFGBV and the UN Special Rapporteur on Freedom of Expression to develop a common definition for gendered disinformation.[125]

Apply a victim-survivor-centred Safety and Privacy by Design approach

Scholarly and policy debates recognise the need to take steps to provide short-term relief and mitigate the risks of OGBV. While user interface design can undermine user agency and safety (e.g., recalling dark patterns outlined above), it can equally enable users to mitigate risks of misogyny.

Reviewing and evaluating immediate responses, research by Sinders, Shukla and Voegeli (2021)[126] and PEN America[127] emphasises the need for platforms to implement improved user tools. Their recommendations propose proactive measures that enable users to reduce risks and exposure to OGBV, reactive measures that facilitate more effective immediate responses when users are faced by OGBV, and accountability measures to aim to deter abusive behaviour and discourage perpetrators from exploiting platforms for networking and coordinated harassment.

A Safety and Privacy by Design approach centres user agency in the development and design of platform products and services. The following measures are not exhaustive, and more research will be needed to evaluate their effectiveness.

Proactive measures may include:

- Content moderation tools such as "shields" that enable users to proactively filter abusive content (across feeds, threads, comments, replies, direct messages, etc.) and quarantine it in a dashboard, where they can review and address it with trusted allies;
- Robust, intuitive, user-friendly features that allow a fine-tuning of privacy and security settings, including "visibility snapshots" that show, in real time, how adjusting settings affects reach;
- Structures that allow users to assemble rapid response teams of trusted allies, including the delegation of account access.

Reactive measures may include:

- Emergency hotlines that users can use to receive trauma-informed support in real time;
- Documentation features that allow users to record evidence of OGBV quickly and easily (for example, instantly capturing screenshots, hyperlinks, and other publicly available data), which should be made interoperable to allow cross-platform evidencing;
- Improved and standardised features to block contacts, mute content, and restrict or hide content;
- Improved reporting mechanisms, including bulk reporting in recognition of coordinated nature of harassment campaigns, as well as circular reporting that allows for a report to be reopened and edited, and across platforms.

Accountability measures may include:

- A transparent system of "escalating penalties" for abusive behaviour, including warnings, strikes, nudges, temporary functionality limitations, suspensions, content takedowns, and account bans. In terms of account bans and de-platforming, research has noted that "removing perpetrators may not get at the root of the problem of accountability," while emphasising that "lock-down mechanisms" should preserve metadata and account information for evidence-gathering and accountability-related purposes;[128]
- Testing "proactive nudges" that aim to encourage users to revise abusive content before they post it (as well as research measuring the efficacy of nudges);
- Sufficiently resourced appeal processes to ensure the clear and time-sensitive review of appeals.

Legislative frameworks have identified the need for user agency and empowerment, and started requiring platforms to be more accountable and transparent about their design and policies. For example, Australia's Online Safety Act 2021 refers to Basic Online Safety Expectations, which require platforms to put in place clear and readily identifiable mechanisms that enable users to report and make complaints about content as well as terms of use, policies, and procedures to deal with complaints and reports. The expectations also require that platforms keep records of user reports and complaints for five years.[129]

Harassment Manager developed by Google's Jigsaw

Harassment Manager is an "open source codebase for a web application that allows users to document and manage abuse targeted at them on social media," starting with X (formerly Twitter), who partnered on the project. The tool intends to help users "identify and document harmful posts, mute or block perpetrators of harassment and hide harassing replies to their own tweets." Users can review tweets based on hashtag, username, keyword or date, leveraging the Perspective API to detect comments that are most likely to be toxic (further discussed below). The Harassment Manager code is available on Github,[130] open sourced for developers and non-governmental organisations to build and adapt for free. This tool should be tested and scrutinised by GBV and feminist advocates and experts as well as victim-survivors to inform the further improvement and development.

Enhance cross-platform cooperation and information sharing

Platforms should recognise that what occurs on other platforms may make its way to their own service (and vice versa). This is not only true of TVEC, but also of the actors and tactics of OGBV. Online harassment campaigns targeting an individual may be coordinated on one platform, with the content or URLs to this content cross-posted to other platforms, where the targeted users may or may not have accounts. As stated earlier, harassment or the coordination of harassment often also involves smaller and 'alternative' platforms.

Platforms should develop and operate exchange channels between relevant teams, including safety and content moderation teams, to proactively share information about cross-platform harassment (such as perpetrators using multiple accounts), including, where relevant, user reports of multi-platform harassment. Platforms should also develop interoperable reporting mechanisms for users to enable user agency and efficient response. Exchange channels may facilitate faster action, for example, when a prominent actor is identified to be linked to repeated harmful behaviour, such as violating community guidelines across platforms. Such efforts are important for understanding the scope and nature of OGBV, but also to coordinate mitigation actions by platforms and other stakeholders, including governments, civil society, or even law enforcement, if relevant.

Already existing cross-platform efforts and crisis protocols such as the Global Internet Forum to Counter Terrorism's Content Incident Protocol[131] and the Christchurch Call Crisis Response Protocol[132] should consider how OGBV is relevant to their scope and mandates, and how to strengthen mechanisms appropriately, recognising misogyny as a radicalisation vector for violent extremism. Relevant voluntary commitments or co-regulatory frameworks could also be reviewed. For example, the EU's 2022 Strengthened Code of Practice on Disinformation includes the commitment to "operate channels of exchange between their relevant teams in order to proactively share information about cross-platform influence operations, foreign interference in information space and relevant incidents that emerge on their respective services, with the aim of preventing dissemination and resurgence on other services."[133] Committed channels could extend to incidents of gendered or sexualised harassment campaigns. Such an effort could also be seen as beneficial to compliance with the EU's DSA under which platforms are required to assess and mitigate systemic risks related to OGBV.

Review and update content moderation policies, processes, and systems

Platforms should assess and mitigate how patriarchal gender norms factor into and are reproduced by their moderation policies and practices. A comprehensive approach to community guidelines and moderation that addresses OGBV, applying a gender lens and a victim-survivor-centred approach, should sensitise policy formulation and enforcement to the continuum of

OGBV as well as the links between misogyny and violent extremism, including the use of dehumanising language based protected characteristics of persons.

Relevant platform teams should consider conducting interviews and focus groups with victims-survivors to inform policy and enforcement processes. Gaps in moderators' lack of understanding of local languages and regional contexts need to be addressed by involving diverse population groups. Civil society has also suggested that platforms provide support, including trauma support, to local organisations who review hate speech policies and develop local lexicons of misogynistic words and phrases.[134]

An example of how AI-based systems could be used to support content moderation is Perspective API,[135] developed by Google's Jigsaw, which uses ML models to identify abusive comments online. Perspective API predicts the perceived impact a comment may have on a conversation by evaluating (scoring) that comment across a range of emotional concepts (attributes). Currently, Perspective API may provide scores for attributes defined as "Toxicity", "Severe Toxicity", "Insult", "Profanity", "Identity attack", "Threat", and "Sexually explicit". The tool intends to help moderators to quickly prioritise and review comments that have been reported, to give feedback to commentators, and for users to control which comments they see. It proposes an encouraging outlook for applying AI-based systems to improve content moderation and to make the online environment safer. Noting that the development and application of such tools is in the early phases, researchers have tested Perspective API to measure levels of toxicity of tweets from prominent drag queens in the US. The research suggest that Perspective considered a significant number of drag queen accounts to have higher levels of toxicity than accounts of white nationalists. Thereby, it was not able to consider the social context when measuring toxicity levels, for example, it did not recognise cases in which words, that might conventionally be seen as offensive, conveyed different meanings in LGBTQ+ speech.[136] This suggests the need for continued and increased multistakeholder collaborations to build on and advance industry tools such as Perspective API.

Both AI-based and human content moderation require comprehensive and regular updates of policies, including trauma-informed processes, to address the nuanced forms of OGBV and prevent counter-productive outcomes. These efforts require a genuine will to improve systems and should not be negatively influenced by company metrics that prioritise engagement. Instead, platform actions should prioritise the principle of 'do no harm', mitigating the exposure to risks.

Audit and mitigate misogyny in AI-based systems

Legislation such as the EU's DSA as well as proposed (non-binding) guidance such as the Violence Against Women and Girls (VAWG) Code of Practice proposed by advocates in the context of the UK's Online Safety Bill (OSB)[137] intend to assess and mitigate the gendered impact of platform design, including their algorithmic feeds.

Articles 34 and 35 of the DSA specifically call on platforms to assess and mitigate the systemic risks posed by their services, including any negative effects in relation to OGBV. Thereby, risk assessments should include the "design of their recommender systems and any other relevant algorithmic system."[138] The VAWG Code emphasises the Safety by Design principle to ensure that "algorithms used on the service do not cause foreseeable harm through promoting hateful content, for example by rewarding misogynistic influencers with greater reach." The Code argues that "preventative measures must consider the role of algorithmic product decisions," reiterating that the decision-making processes around their development and deployment must be scrutinised.[139] Algorithmic accountability and auditing should take a victim-survivor-centred approach and conduct safety testing, and apply gender analysis and intersectional perspectives, specifically testing how individual users experience intersecting forms of identity-based hate and violence.

Auditing and evaluating the impact of algorithms remains a challenge, even with direct access to proprietary code, given that algorithmic feeds are personalised and rely on many factors including users' historical data. Moreover, independent auditors need to use a counterfactual scenario to compare the algorithmic feed, for example, when conducting randomised controlled experiments. For example, a recent study assigned a sample of consenting users to reverse-chronologically-ordered feeds to assess the impact of algorithmic feeds, including

how they encourage partisan stereotyping or influence negative attitudes about outgroups.[140] In the context of OGBV, studying the role of algorithmic recommendations in the Manosphere may help with developing evidence-based interventions in online radicalisation pathways, given that it is likely that users who enter the Manosphere may have less intense, less extreme beliefs and slowly form new connections and become further embedded within the inner community.[141]

Christchurch Call Initiative on Algorithmic Outcomes

The Christchurch Call Initiative on Algorithmic Outcomes, led by New Zealand, the US, X (formerly Twitter) and Microsoft, seeks to develop software tools to facilitate independent research on the impact of user interactions with algorithmic systems.[142] Working with OpenMined, DailyMotion and LinkedIn, a new software infrastructure will integrate privacy enhancing technologies to allow external researchers and data scientists to remotely study algorithms distributed across multiple secure sites. Such effort is crucial to enable independent research on the impact of algorithmic feeds. The independent auditing of algorithms and ML models via such software infrastructure should focus on understanding and testing the role of algorithmic and pathways, including, where possible, across platforms. The development of systems for remote researcher queries will need appropriate governance and ethics frameworks as well as processes for research prioritisation.

In terms of potential mitigation of biased AI systems, scholars have suggested that ML models can be developed so that they do not produce discriminatory patterns such as gender stereotypes. The idea would be not to limit the data input (i.e., remove any data related to gender), but to prevent algorithms from yielding gender-based patterns, since not using gender data may still allow for predicting gender and result in discrimination by proxy.[143] For example, risk mitigation could involve interventions for bias reduction, including debiasing an algorithm's training set.[144] Transparency and inclusivity, by incorporating intersectional feminist knowledge, will be critical for algorithmic auditing.

Conclusion

Normalising misogynistic violence, the harassment and intimidation of women and the LGBTQ+ community, and upholding patriarchal gender norms, are all situated in "larger patterns of systemic violence made to control, demean, and significantly limit the autonomy of the person targeted."[145] Online manifestations of GBV impede the safety, freedom of expression, and participation in public life of women and LGBTQ+ people.

In this context, the paper has emphasised the continuum of violence within which misogyny can serve as an ideological link across different forms of violent extremism. A recognition of the systemic issue of patriarchal norms in society and the risk of misogyny as an ideology that can be a gateway to radicalisation should be reflected in a systemic response by platforms, governments, and civil society.

Beyond immediate action to enable victim-survivor-centred user agency, platforms should assess both individual and societal level harm caused by OGBV, especially recognising the need to consider the relationship between misogyny and TVEC in their community guidelines and risk assessments. Evidence gaps in the research on violent extremism and misogyny reiterate the need to further study these complex phenomena, including by means of strengthening data access and transparency reporting. Evidence-based decision-making and interventions are central to avoid any potential negative consequences for achieving gender equality and safeguarding freedom of expression.

Platforms should develop inclusive community guidelines and sufficiently invest in clear and consistent enforcement. Platforms should also assess and mitigate any risks stemming from the functioning of their systems, which include algorithmic recommender systems. Notably, product teams should assess how algorithmic design (and unintended consequences) reflect and reproduce patriarchal gender norms that risk amplifying misogynistic content. A Safety by Design approach should ensure inclusive design and safety testing that incorporates intersectional perspectives before launching new services and features. Finally, meaningful access to platform data via APIs for vetted researchers remains fundamental to gathering and understanding evidence, making sense of research findings, and holding platforms accountable. This paper and its recommendations should be understood as complementing whole-of-society and whole-of-government actions as platforms can play a crucial role in supporting such efforts.

Endnotes

1 Public Service Commission of New Zealand. (n.d.). Diversity and inclusion glossary. Retrieved from https://www.publicservice.govt.nz/guidance/glossary/diversity-and-inclusion/

2 Office of the United Nations High Commissioner for Human Rights (OHCHR). (n.d.). Gender stereotyping. Retrieved from https://www.ohchr.org/en/women/gender-stereotyping

3 European Commission. (n.d.). Gender-based violence. Retrieved from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence_en

4 Southern Poverty Law Center. (n.d.). Male supremacy. Retrieved from https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy

5 Connell, R. (1987). Gender and Power. Sydney: Allen and Unwin.

6 Institute for Strategic Dialogue (ISD). (n.d.). The 'Manosphere'. Retrieved from https://www.isdglobal.org/explainers/the-manosphere-explainer/

7 Manne, K. (2019). Down girl: the logic of misogyny. New York: Oxford University Press.

8 Gentry, C. E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Retrieved from https://doi.org/10.1080/17539153.2022.2031131

9 Bailey, M., & Trudy. (2018). On Misogynoir: Citation, Erasure, and Plagiarism. Feminist Media Studies, 18(4), 1–7. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/14680777.2018.1447395?journalCode=rfms20

10 Gentry, C.E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Retrieved from https://doi.org/10.1080/17539153.2022.2031131

11 UN Women. (2023). Expert Group Meeting report: Technology-facilitated violence against women: Towards a common definition. Retrieved from https://www.unwomen.org/en/digital-library/publications/2023/03/expert-group-meeting-report-technology-facilitated-violence-against-women

12 Manne, K. (2019). Down girl: the logic of misogyny. New York: Oxford University Press.

13 United Nations Office of the High Commissioner for Human Rights. (2018). The impact of online violence on women human rights defenders and women's organisations. Retrieved from https://www.ohchr.org/en/statements/2018/06/impact-online-violence-women-human-rights-defenders-and-womens-organisations?LangID=E&NewsID=23238

14 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Retrieved from https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics

15 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Retrieved from https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists

16 Di Meco, L. (2023). Monetizing Misogyny. ShePersisted. Retrieved from https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted_MonetizingMisogyny.pdf

17 United Nations Digital Library. (2018). Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective. A/HRC/38/47. Retrieved from https://digitallibrary.un.org/record/1641160

18 Gentry, C. E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Retrieved from https://doi.org/10.1080/17539153.2022.2031131

19 Georgetown Institute for Women, Peace and Security. (n.d.). Recognizing the violent extremist ideology of incels. Retrieved from https://giwps.georgetown.edu/resource/recognizing-the-violent-extremist-ideology-of-incels/

20 U.S. Department of State. (2022). 2022 roadmap for the global partnership for action on gender-based online harassment and abuse. Retrieved from https://www.state.gov/2022-roadmap-for-the-global-partnership-for-action-on-gender-based-online-harassment-and-abuse/

21 APC Women's Rights Programme. (2015). Briefing paper on VAW. Retrieved from https://www.apc.org/sites/default/files/HRC%2029%20VAW%20a%20briefing%20paper_FINAL_June%202015.pdf

22 Gentry, C. E. (2022). Misogynistic Terrorism: It Has Always Been Here. Critical Studies on Terrorism 15: 209–224. Retrieved from https://www.tandfonline.com/doi/full/10.1080/17539153.2022.203113; Johnston, M. F., M. Iqbal, & J. True. (2020). The Lure of (Violent) Extremism: Gender Constructs in Online Recruitment and Messaging in Indonesia. Studies in Conflict & Terrorism 0731 (May): 1–19. Retrieved from https://doi.org/10.1080/1057610X.2020.1759267; Phelan, A. (2021). Terrorism, gender and women: toward an integrated research agenda. Retrieved from https://www.routledge.com/Terrorism-Gender-and-Women-Toward-an-Integrated-Research-Agenda/Phelan/p/book/9780367623104; Meiering, D., A. Dziri, & N. Foroutan. (2020). Connecting Structures: Resistance, Heroic Masculinity and Anti-Feminism as Bridging Narratives within Group Radicalization." International Journal of Conflict and Violence (IJCV) 14 (2): 1–19. Retrieved from https://doi.org/10.4119/IJCV-3805

23 Phelan, A., White, J., Wallner, C., & Paterson, J. (2023). Hostile Beliefs And The Transmission Of Extremism: A Comparison Of The Far-Right In The UK And Australia. Centre for Research and Evidence on Security Threats (CREST). Retrieved from https://crestresearch.ac.uk/resources/misogyny-hostile-beliefs-and-the-transmission-of-extremism/

24 Institute for Strategic Dialogue (ISD). (n.d.). The Manosphere. Retrieved from https://www.isdglobal.org/explainers/the-manosphere-explainer/

25 Southern Poverty Law Center. (n.d.). Male supremacy. Retrieved from https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy

26 Ibid.

27 Institute for Strategic Dialogue (ISD). (n.d.). Incels. Retrieved from https://www.isdglobal.org/explainers/incels/

28 Gheorghe, R. M. (2023). "Just Be White (JBW)": Incels, Race and the Violence of Whiteness. Feminist Inquiry in Social Work, 1-19. Retrieved from https://doi.org/10.1177/08861099221144275

29 Institute for Strategic Dialogue (ISD). (n.d.). Incels. Retrieved from https://www.isdglobal.org/explainers/incels/

30 For example: Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. Studies in Conflict & Terrorism, 43(7), 565-587. Retrieved from https://doi.org/10.1080/1057610X.2020.1751459

31 Public Prosecution Service of Canada. (2023). Court rules that murder and attempted murder were terrorist activity in youth case. Retrieved from https://www.ppsc-sppc.gc.ca/eng/nws-nvs/2023/27_07_23.html

32 Davey, D., Comerford, M., Guhl, J., Baldet, W., & Colliver, C. (2021). A taxonomy for the classification of post-organisational violent terrorist content. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/a-taxonomy-for-the-classification-of-post-organisational-violent-terrorist-content/

33 Fielitz, M., Ebner, J., Guhl, J., & Quent, M. (2018). Loving Hate: Anti-Muslim Extremism, Radical Islamism and the Spiral of Polarization. Institute for Democracy and Civil Society (IDZ) / Institute for Strategic Dialogue (ISD). Retrieved from https://www.idz-jena.de/fileadmin/user_upload/antimuslimextremism_radicalislamism_polarization.pdf

34 Cook, J. & Roose, J. (2023). Supreme Men, Subjugated Women: Gender Inequality and Violence in Jihadist, Far-Right, and Male Supremacist Ideologies. GNET Research. Retrieved from https://gnet-research.org/2023/01/10/supreme-men-subjected-women-gender-inequality-and-violence-in-jihadist-far-right-and-male-supremacist-ideologies/

35 Ibid.

36 Antisemitism Policy Trust (2019). Misogyny and Antisemitism: Briefing. Retrieved from https://antisemitism.org.uk/wp-content/uploads/2019/05/5982-Misogyny-and-Antisemitism-Briefing-April-2019-v1.pdf

37 Crawford, C. (2022). Sleeping with the Enemy: Sex, Sexuality and Antisemitism in the Extreme Right. Institute for Community and Social Research. Retrieved from https://icsr.info/2022/06/22/sleeping-with-the-enemy-sex-sexuality-and-antisemitism-in-the-extreme-right/

38 Anti-Defamation League. (2023). Antisemitism & Anti-LGBTQ+ Hate Converge in Extremist and Conspiratorial Beliefs. Center on Extremism. Retrieved from https://www.adl.org/resources/blog/antisemitism-anti-lgbtq-hate-converge-extremist-and-conspiratorial-beliefs

39 Anti-Defamation League. (2022) What is "Grooming?" The Truth Behind the Dangerous, Bigoted Lie Targeting the LGBTQ+ Community. Center on Extremism. Retrieved from https://www.adl.org/resources/blog/what-grooming-truth-behind-dangerous-bigoted-lie-targeting-lgbtq-community

40 Anti-Defamation League. (2023). Antisemitism & Anti-LGBTQ+ Hate Converge in Extremist and Conspiratorial Beliefs. Center on Extremism. Retrieved from https://www.adl.org/resources/blog/antisemitism-anti-lgbtq-hate-converge-extremist-and-conspiratorial-beliefs

41 Sykes, S. &. Dr Hopner, V. (2023). Tradwives: The Housewives Commodifying Right-Wing Ideology. GNET Research. Retrieved from https://gnet-research.org/2023/07/07/tradwives-the-housewives-commodifying-right-wing-ideology/

42 Veilleux-Lepage, Y. et al. (2023). Gendered radicalisation and 'everyday practices': An analysis of extreme right and Islamic State women-only forums. European Journal of International Security (2023), 8, 227–242. Retrieved from https://researchmgt.monash.edu/ws/portalfiles/portal/461281934/420784008_oa.pdf

43 Chan, E. (2023). Technology-Facilitated Gender-Based Violence, Hate Speech, and Terrorism: A Risk Assessment on the Rise of the Incel Rebellion in Canada. Violence Against Women. Vol. 29, Issue 9, 1687-1718. Retrieved from https://doi.org/10.1177/10778012221125495

44 Martiny, C. & Lawrence, S. (2023). A year of hate: Anti-drag mobilization efforts targeting LGBTQ people in the US. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/a-year-of-hate-anti-drag-mobilization-efforts-targeting-lgbtq-people-in-the-us/

45 Argentino, M.-A., Raja, A. & Gallagher, A. (2022). She Drops: How QAnon Conspiracy Theories Legitimize Coordinated and Targeted Gender Based Violence. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/she-drops-how-qanon-conspiracy-theories-legitimize-coordinated-and-targeted-gender-based-violence/

46 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Forthcoming report]

47 Kelly, L. (1988). Surviving sexual violence. Cambridge, UK: Polity Press.

48 Dunn, S. (2021). Is it Actually Violence? Framing Technology-Facilitated Abuse as Violence. In: Bailey, J., Flynn, A. & Henry, N. (eds) The Emerald International Handbook of Technology-Facilitated Violence and Abuse (Bingley, UK: Emerald Publishing, 2021). Retrieved from https://doi.org/10.1108/978-1-83982-848-520211002

49 The Economist Intelligence Unit (EIU). (2021). Measuring the prevalence of online violence against women. Retrieved from https://onlineviolencewomen.eiu.com/

50 PEN America. (n.d.). Defining online harassment: A glossary of terms. Retrieved from https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms/#cross

51 The Global Partnership. (2023). Technology-facilitated gender-based violence: Preliminary landscape analysis. The Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership). Retrieved from https://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis

52 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Retrieved from https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists

53 Jankowicz, N. et al. (2021). Malign creativity: How gender, sex, and lies are weaponized against women online. Wilson Center. Science and Technology Innovation Program. Retrieved from https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online

54 Lenhart, A., Ybarra, M., Zickuhr, K. & Price-Feeney, M. (2016). Online Harassment, Digital Abuse, and Cyberstalking in America. Data & Society Research Institute. Retrieved from https://datasociety.net/library/online-harassment-digital-abuse-cyberstalking/

55 Argentino, M.-A., Raja, A. & Gallagher, A. (2022). She Drops: How QAnon Conspiracy Theories Legitimize Coordinated and Targeted Gender Based Violence. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/she-drops-how-qanon-conspiracy-theories-legitimize-coordinated-and-targeted-gender-based-violence/

56 Dunn, S. (2020). Technology-Facilitated Gender-Based Violence: An Overview. Supporting a Safer Internet Paper No. 1. Centre for International Governance Innovation (CIGI). Retrieved from https://www.cigionline.org/publications/technology-facilitated-gender-based-violence-overview/

57 The Global Partnership. (2023). Technology-facilitated gender-based violence: Preliminary landscape analysis. The Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership). Retrieved from https://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis

58 Ibid.

59 Ibid.

60 PEN America. (n.d.). Defining online harassment: A glossary of terms. Retrieved from https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms/#cross

61 United Nations Digital Library. (2018). Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective. A/HRC/38/47. Retrieved from https://digitallibrary.un.org/record/1641160

62 Dunn, S. (2020). Technology-Facilitated Gender-Based Violence: An Overview. Supporting a Safer Internet Paper No. 1. Centre for International Governance Innovation (CIGI). Retrieved from https://www.cigionline.org/publications/technology-facilitated-gender-based-violence-overview/

63 Henry, N. et al. (2020). Image-based Sexual Abuse. A Study on the Causes and Consequences of Non-consensual Nude or Sexual Imagery. Routledge. Retrieved from https://doi.org/10.4324/9781351135153

64 Wittes, B., Poplin, C., Jurecic, Q. & Spera, C., (2016). Sextortion: Cybersecurity, teenagers, and remote sexual assault. Brookings Institute. Retrieved from https://www.brookings.edu/articles/sextortion-cybersecurity-teenagers-and-remote-sexual-assault/

65 Palmer, T. (2018). Rape pornography, cultural harm and criminalization. Northern Ireland Legal Quarterly. Vol. 69 No. 1 (2018): Spring. Retrieved from https://doi.org/10.53386/nilq.v69i1.77

66 Viola, M., & Voto, C. (2023). Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. Synthese. 201. Article number: 30. Retrieved from https://doi.org/10.1007/s11229-022-04012-2

67 Citron, D. (2019). Sexual Privacy. Yale Law Journal 128 (7): 1870–1960.

68 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Retrieved from https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists

69 Veletsianos, G., Houlden, S., Hodson, J. & Gosse, C. (2018). Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. New Media & Society 20 (12): 4689–708. Retrieved from

70 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Forthcoming report]

71 United Nations Office of the High Commissioner for Human Rights. (2018). The impact of online violence on women human rights defenders and women's organisations. Retrieved from https://www.ohchr.org/en/statements/2018/06/impact-online-violence-women-human-rights-defenders-and-womens-organisations?LangID=E&NewsID=23238

72 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Retrieved from https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists

73 Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. Studies in Conflict & Terrorism, 43(7), 565-587. Retrieved from https://doi.org/10.1080/1057610X.2020.1751459

74 Simmons, C. & Fourel, Z. (2022). Hate in Plain Sight: Abuse Targeting Women Ahead of the 2022 Midterm Elections on TikTok & Instagram. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/hate-in-plain-sight-abuse-targeting-women-ahead-of-the-2022-midterm-elections-on-tiktok-instagram/

75 Binder, J. F., & Kenyon, J. (2022). Terrorism and the internet: How dangerous is online radicalization?. Frontiers in psychology, 6639. Retrieved from https://doi.org/10.3389/fpsyg.2022.997390

76 Bates, S. (2016). Revenge porn and mental health: A qualitative analysis of the mental health effects of revenge porn on female survivors. Feminist Criminology, 12(1), 22.

77 Martiny, M., Visser, F., & Jones, I. (2022). Evaluating Platform Abortion-Related Speech Policies: Were Platforms Prepared for the Post-Dobbs Environment? Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/evaluating-platform-abortion-related-speech-policies-were-platforms-prepared-for-the-post-dobbs-environment/

78 Jane, E. (2018). Gendered cyberhate as workplace harassment and economic vandalism. Feminist Media Studies, 18(4), 575–591.

79 UN Women. (2023). Brief: The state of evidence and data collection on technology-facilitated violence against women. Retrieved from https://www.unwomen.org/en/digital-library/publications/2023/04/brief-the-state-of-evidence-and-data-collection-on-technology-facilitated-violence-against-women

80 Plan International. (2022). State of the World's Girls Report 2021: Communications Report. Retrieved from https://plan-international.org/uploads/2022/02/sotwgr2021-commsreport-en.pdf

81 UNFPA. (n.d.). Preventing Technology-Facilitated Gender-Based Violence (TF GBV). Retrieved from https://www.un.org/techenvoy/sites/www.un.org.techenvoy/files/GDC-Submission_UNFPA.pdf

82 U.S. Secret Service Media Relations. (2023). New Secret Service Research Examines for the First Time Five Years of Mass Violence Data. Retrieved from https://www.secretservice.gov/newsroom/releases/2023/01/new-secret-service-research-examines-first-time-five-years-mass-violence

83 U.S. Department of State. (2023). Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse. Retrieved from https://www.state.gov/2023-roadmap-for-the-global-partnership-for-action-on-gender-based-online-harassment-and-abuse/

84 She Persisted. (2023). Monetizing Misogyny. Retrieved from https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted_MonetizingMisogyny.pdf

85 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Retrieved from https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists

86 Global Engagement Center. (2023). Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors. Retrieved from https://www.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/

87 ISD research reviewed the policies of X (formerly Twitter), Meta's Facebook and Instagram, YouTube, TikTok and Telegram.

88 Martiny, M., Visser, F., & Jones, I. (2022) Evaluating Platform Abortion-Related Speech Policies: Were Platforms Prepared for the Post-Dobbs Environment? Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/evaluating-platform-abortion-related-speech-policies-were-platforms-prepared-for-the-post-dobbs-environment/

89 Oversight Board. (2023). Oversight Board Overturns Meta's Original Decision in Image of Gender-Based Violence Case. Retrieved from https://oversightboard.com/news/813577783586004-oversight-board-overturns-meta-s-original-decision-in-image-of-gender-based-violence-case/

90 X Help Center (2023). Hateful Conduct. Retrieved from https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy

91 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Forthcoming report]

92 Ibid.

93 O'Connor, C. (2021). Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/hatescape-an-in-depth-analysis-of-extremism-and-hate-speech-on-tiktok/

94 Wired UK. (2022). Telegram Has a Serious Doxing Problem. Retrieved from https://www.wired.co.uk/article/telegrams-doxing-problem

95 Holnburger, J. (2023. Chronology of a radicalization: How Telegram became the most important platform for conspiracy ideologies and right-wing extremism. CeMAS. Retrieved from https://cemas.io/en/publications/chronology-of-a-radicalization/

96 Gerster, L., Kuchta, R., Hammer, D. & Schwieter, C. (2022). Telegram as a Buttress: How Far-Right Extremists and Conspiracy Theorists Are Expanding Their Infrastructures via Telegram. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/telegram-as-a-buttress-how-far-right-extremists-and-conspiracy-theorists-are-expanding-their-infrastructures-via-telegram/

97 Council for Media Services in Slovakia (CMS) & Reset (2023). The Bratislava Shooting Report on the role of online platforms. Retrieved from https://rpms.sk/sites/default/files/2023-03/CMS_RESET_Report.pdf

98 Deceptive Design. (n.d.). Types of deceptive pattern. Retrieved from https://www.deceptive.design/types

99 Sinders, C., Shukla, V. & Voegeli, E. (2021). Trust Through Trickery. Common Place. Retrieved from https://commonplace.knowledgefutures.org/pub/trust-through-trickery/release/1

100 Ibid.

101 Ibid.

102 eSafety Commissioner. (n.d.). Safety by Design. Retrieved from http://www.esafety.gov.au/industry/safety-by-design

103 Ada Lovelace Institute. (2021). Technical methods for regulatory inspection of algorithmic systems. Retrieved from https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/

104 Integrity Institute. (n.d.). Ranking by Engagement. Retrieved from https://integrityinstitute.org/blog/ranking-by-engagement

105 Thomas, E. & Balint, K. (2022). Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the Manosphere. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere/

106 Simmons, C. & Fourel, Z. (2022). Hate in Plain Sight: Abuse Targeting Women Ahead of the 2022 Midterm Elections on TikTok & Instagram. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/hate-in-plain-sight-abuse-targeting-women-ahead-of-the-2022-midterm-elections-on-tiktok-instagram/

107 Sykes, S. &. Dr Hopner, V. (2023). Tradwives: The Housewives Commodifying Right-Wing Ideology. GNET Research. Retrieved from https://gnet-research.org/2023/07/07/tradwives-the-housewives-commodifying-right-wing-ideology/

108 Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

109 Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. ProPublica. Retrieved from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

110 Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366, 447–453.

111 Vlasceanu, M. & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. PNAS. Vol. 119, No. 29. Retrieved from https://www.pnas.org/doi/full/10.1073/pnas.2204529119#body-ref-r2-2

112 Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the First Conference on Fairness, Accountability and Transparency, 77–91. PMLR. Retrieved from https://proceedings.mlr.press/v81/buolamwini18a.html

113 Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021). Gendering Algorithms in Social Media. ACM SIGKDD Explorations Newsletter, 23(1), 24–31. Retrieved from https://doi.org/10.1145/3468507.3468512

114 Scott, M. (2023) We Don't Talk About Fixing Social Media. POLITICO. Retrieved from https://www.politico.eu/newsletter/digital-bridge/we-dont-talk-about-fixing-social-media/

115 It is important to consider that the type of data deemed "publicly accessible" can change. Thereby, tracking any changes should build on a shared taxonomy of what data is "publicly accessible."

116 Public interest research commonly refers to research with the explicit aim to develop society's collective knowledge. Regulatory precedent suggests that public interest research must be independent of commercial interests and reveal the source of its funding. Public interest researchers are not necessarily linked to academic institutions and can also include researchers affiliated to non-profit or media organisations.

117 Bundtzen, S. & Schwieter, C. (2023). Researcher Access to Social Media Data: Lessons Learnt & Recommendations for Strengthening Initiatives in the EU & Beyond. Institute for Strategic Dialogue (ISD). Retrieved from https://www.isdglobal.org/isd-publications/researcher-access-to-social-media-data-lessons-learnt-recommendations-for-strengthening-initiatives-in-the-eu-beyond/

118 Wilton Park. (2022). Building a Shared Agenda on the Evidence Base for Gender-Based Online Harassment and Abuse. WP3057. Retrieved from https://www.wiltonpark.org.uk/event/building-a-shared-agenda-on-the-evidence-base-for-gender-based-online-harassment-and-abuse/

119 UN Women. (2023). Technology-Facilitated Violence Against Women: Taking Stock of Evidence and Data Collection. Retrieved from https://www.unwomen.org/en/digital-library/publications/2023/04/technology-facilitated-violence-against-women-taking-stock-of-evidence-and-data-collection

120 Methodologies include survey data generated through household, population-based, or experimental and specialised online surveys; quantitative administrative data, including service-based, programmatic or transparency reporting data; qualitative data, generated by key informant interviews and focus group discussions, or digital ethnographies; as well as data generated through mixed-methods, including by means of machine learning to collect and analyse social media data.

121 Conway, M. (2016). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. Studies In Conflict & Terrorism. 2017, Vol. 40, No. 1, 77–98. Retrieved from https://www.tandfonline.com/doi/full/10.1080/1057610X.2016.1157408

122 European Commission. (2022). 2022 Strengthened Code of Practice on Disinformation. Retrieved from https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

123 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Retrieved from https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics

124 UN Women. (2023). Technology-Facilitated Violence Against Women: Taking Stock of Evidence and Data Collection. Retrieved from https://www.unwomen.org/en/digital-library/publications/2023/04/technology-facilitated-violence-against-women-taking-stock-of-evidence-and-data-collection

125 Office of the United Nations High Commissioner for Human Rights (OHCHR). (2023). Report on freedom of expression and the gender dimensions of disinformation. Retrieved from https://www.ohchr.org/en/calls-for-input/2023/report-freedom-expression-and-gender-dimensions-disinformation

126 Sinders, C., Shukla, V. & Voegeli, E. (2021). Trust Through Trickery. Common Place. Retrieved from https://commonplace.knowledgefutures.org/pub/trust-through-trickery/release/1

127 PEN America. (n.d.). No Excuse for Abuse. What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users. Retrieved from https://pen.org/report/no-excuse-for-abuse/

128 Anti-Defamation League (ADL). (2018). When Women Are the Enemy: Intersection Misogyny and White Supremacy. Retrieved from https://www.adl.org/resources/report/when-women-are-enemy-intersection-misogyny-and-white-supremacy

129 eSafety Commissioner. (n.d.). Basic Online Safety Expectations. Retrieved from https://www.esafety.gov.au/industry/basic-online-safety-expectations

130 GitHub. (n.d.). Harassment Manager. Retrieved from https://github.com/conversationai/harassment-manager

131 Global Internet Forum to Counter Terrorism (GIFCT). (n.d.). Content Incident Protocol. Retrieved from https://gifct.org/content-incident-protocol/

132 Christchurch Call. (2022). Christchurch Call Initiative on Algorithmic Outcomes. Retrieved from https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/

133 European Commission. (2022). 2022 Strengthened Code of Practice on Disinformation. Retrieved from https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

134 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Retrieved from https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics

135 Perspective (n.d.) Developers. About the API. Retrieved from https://developers.perspectiveapi.com/s/about-the-api?language=en_US

136 Dias Oliva, T., Antonialli, D.M. & Gomes, A. (2021). Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. Sexuality & Culture 25, 700–732. Retrieved from: https://doi.org/10.1007/s12119-020-09790-w

137 The End Violence Against Women Coalition, Glitch, Refuge, Carnegie UK, NSPCC, 5Rights, Professor Clare McGlynn and Professor Lorna Woods (2022). Violence Against Women and Girls (VAWG) Code of Practice. Retrieved from https://carnegieuktrust.org.uk/publications/violence-against-women-and-girls-vawg-code-of-practice/

138 Official Journal of the EU (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Volume 65. 27 October. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065&from=EN

139 Council of Europe. (2017). Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Retrieved from https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html

140 Andrew M. Guess et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign?. Science 381,398-404. Retrieved from https://www.science.org/doi/10.1126/science.abp9364

141 Habib, H., et al. (2022). Making a Radical Misogynist: How Online Social Engagement with the Manosphere Influences Traits of Radicalization. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), Article No. 450, 1–28. Retrieved from https://dl.acm.org/doi/10.1145/3555551

142 Christchurch Call. (2022). Christchurch Call Initiative on Algorithmic Outcomes. Retrieved from https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/

143 Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021). Gendering Algorithms in Social Media. ACM SIGKDD Explorations Newsletter, 23(1), 24–31. Retrieved from https://doi.org/10.1145/3468507.3468512

144 Gebru, T., et al. (2022). Excerpt from datasheets for datasets. In Ethics of Data and Analytics, K. Martin (Ed.), Auerbach Publications, 148–156. Retrieved from

145 Dunn, S. (2021). Is it Actually Violence? Framing Technology-Facilitated Abuse as Violence. In: Bailey, J., Flynn, A. & Henry, N. (eds) The Emerald International Handbook of Technology-Facilitated Violence and Abuse (Bingley, UK: Emerald Publishing, 2021). Retrieved from https://doi.org/10.1108/978-1-83982-848-520211002

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2023). The Institute for Strategic Dialogue (gGmbH) is registered with the Local Court of Berlin-Charlottenburg (HRB 207 328B). The Executive Director is Huberta von Voss. The address is: PO Box 80647, 10006 Berlin. All rights reserved.

www.isdgermany.org

Sponsored by:

Federal Foreign Office