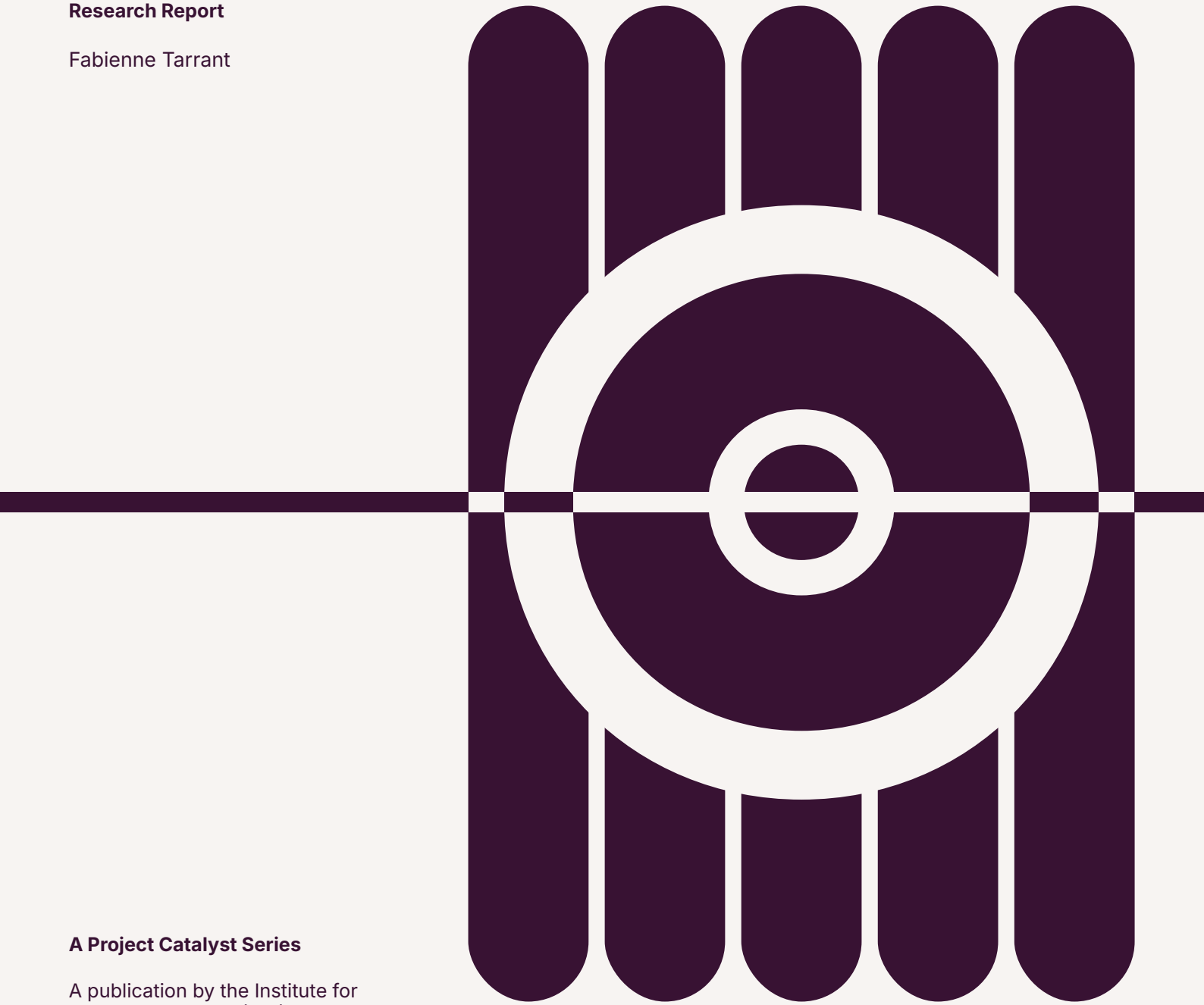
05

Beyond the Silos: Platform Policies on Misogyny, TFGBV and Violent Extremism

October 2026

Research Report

Fabienne Tarrant



A Project Catalyst Series

A publication by the Institute for Strategic Dialogue (ISD), produced as part of Project Catalyst led by the Christchurch Call Foundation.

ISD | Institute for Strategic Dialogue

Christchurch Call Foundation

Research Report

Beyond the Silos: Platform Policies on Misogyny, TFGBV and Violent Extremism

Executive Summary

Key Findings 5

Introduction 6

How to read this report 8

Glossary 9

Methodology

Scope 11

Analysis criteria 12

Limitations 13

The Overall Policy Landscape for Gender-Based Violence Harms

Table 1: TFGBV Harm Coverage Across 12 Platforms 15

Where harms are located 17

Gender dimension 17

Which harms are better covered 18

Violent Extremism and Dangerous Organizations, Hate and Violence Policies: Coverage of TFGBV Harms

Table 2: Violent Extremism and Dangerous Organizations, Hate, & Violence Policies 20

Where harms are located 23

Gender dimension 23

Which harms are captured and how 24

Additional Policies 24

Cross-cutting themes 26

Recommendations 28

Conclusion 31

Annex: Policy Inventory Table 32

Copyright © Institute for Strategic Dialogue (2026)

A Project Catalyst Series

This report is part of Project Catalyst[1], a series by the Institute for Strategic Dialogue (ISD) examining the links between misogyny, targeted hate and violent extremism online. The series includes ISD's practical toolkit for addressing these intersections, companion reports mapping online ecosystems in Canada and Jordan, and a report on how platforms address this intersection in their policies.

1 Project Catalyst is a scalable 18-month consortium project tackling online misogyny and gender-based violence as pathways to radicalization and violent extremism. Led by the Christchurch Call Foundation with implementing partners including ISD, Meedan, Search for Common Ground, PERIL at American University, and the Blavatnik School of Government at Oxford, Catalyst delivers prevention programs across Canada, Jordan, Kenya, and transnational gaming communities.

Executive Summary

In this section

Key Findings	**5**
Introduction	**6**
How to read this report	**8**
Glossary	**9**

As part of Project Catalyst, the Institute for Strategic Dialogue (ISD) assessed 120 published policy documents across 12 platforms to examine how they address technology-facilitated gender-based violence (TFGBV) and its intersections with targeted hate and violent extremism. The analysis finds that while all platforms prohibit at least some forms of TFGBV-related behavior, few consistently recognize its gendered dimensions or account for how misogynistic harm intersects with extremism. Significant gaps remain in the recognition of male-supremacist extremism, the scope of protected characteristics and protections for adults affected by sexual extortion. However, examples across existing policies demonstrate opportunities to better integrate gender-based harms into platform governance. The report identifies practical steps platforms can take to strengthen these protections and address gaps across policy areas.

Key Findings

Finding 1

All 12 platforms prohibit at least some forms of TFGBV-related behavior, but recognition of its gendered dimensions is uneven.

Harassment is the only harm addressed by every platform, followed by intimate image abuse and sexual extortion. Online impersonation and account access control are addressed less consistently and, where covered, are generally framed as gender-neutral security or authenticity issues.

Finding 2

Platform protections do not consistently reflect the diversity of gender-related characteristics.

No platform's protected characteristics list includes gender expression, sex characteristics or non-adherence to prescribed gender roles or norms. Four platforms (Rumble, Telegram, 4chan and Incels.is) do not list any protected characteristics.

Finding 3

Violent extremism policies rarely account for the gendered nature of TFGBV.

Where gender is recognized, these policies largely rely on protected characteristics defined in separate hate policies. Harassment is the only TFGBV category they fully cover; other policies address conduct that overlaps with TFGBV tactics without considering who is targeted.

Finding 4

Male-supremacist extremism is absent from platforms' violent extremism and dangerous organizations policies.

This contrasts with the explicit naming of other supremacist ideologies in some policies. For example, Meta's Dangerous Organizations and Individuals policy lists Nazism, white supremacy, white separatism and white nationalism. TikTok also identifies misogyny as a supremacist ideology in its hate policy, but not in its policy dedicated to violent extremism.

Finding 5

Protections can vary depending on users' jurisdictions.

TikTok's US Terms of Service do not include a harassment prohibition, unlike its Terms of Service for the European Economic Area, United Kingdom and Switzerland, as well as other regions outside the US.

Finding 6

Sexual extortion is frequently framed as a child-safety issue, leaving gaps in protections for adults.

X's only coverage of sexual extortion appears in its Child Safety policy, while YouTube's sole, partial coverage is also focused on child safety.

Finding 7

Some platforms already demonstrate ways to bridge policy silos.

Meta's Violence and Incitement policy is the only violence policy analyzed that explicitly prohibits the glorification of gender-based violence. Its Coordinating Harm and Promoting Crime policy also recognizes the outing of LGBTQI+ people and unveiled women as a harm in itself. Twitch's Off-Service Conduct policy places sexual assault and grooming alongside violent extremism and hate group membership as grounds for the same enforcement outcome.

Introduction

This report is part of Project Catalyst, a series by the Institute for Strategic Dialogue (ISD) examining the intersections of misogyny, targeted hate and violent extremism online. The series includes ISD's practical Toolkit for addressing these intersections, companion reports examining online ecosystems in Canada and Jordan, and a playbook for designing and implementing positive interventions in gaming communities.

Technology-facilitated gender-based violence (TFGBV), targeted hate and violent extremism can overlap in ways that cut across multiple areas of platform governance. Misogynistic abuse, for example, may fall under policies addressing harassment and bullying, hate speech, sexual content, privacy and security, violence, or violent extremism and dangerous organizations. This is dependent upon how it manifests and who or what it targets. These policy areas are often developed separately, with distinct definitions, enforcement thresholds and protected characteristics. As a result, harms that cut across policy areas risk being understood or addressed in isolation. To complicate matters further, extremism and targeted hate usually also have their own policy area, that is not always gendered. Understanding how platforms address the intersections between these areas is therefore essential to identifying how gender-based harms are recognized and addressed.

To examine how platforms navigate these intersections, ISD assessed 120 published policy documents across 12 platforms. These platforms were selected based on those examined in the accompanying Project Catalyst research on how these harms manifest in online ecosystems in Canada and Jordan. The analysis uses the five categories of TFGBV identified in ISD's Toolkit: online harassment, intimate image abuse, sexual extortion, account access control and online impersonation. It examines where these harms are addressed across public platform policies, whether their gendered dimensions are explicitly recognized, and how policies account for the intersection of misogynistic and gender-based harms with targeted hate, violence and violent extremism.

This report approaches misogyny and extremism as interconnected phenomena: misogyny can function as a standalone violent extremist ideology, form a core component of other supremacist ideologies, and act as both a risk factor and warning sign of violent extremism. It also understands TFGBV, targeted hate and violent extremism as existing along a shared continuum of harms that can exclude women and LGBTQI+ people from digital and public spaces, undermining their equal participation in social, political and democratic life.

NOTE **Conceptual note: How this report understands Misogyny, TFGBV and violent extremism**

This report draws on elements of the Toolkit's framing that underpin the analysis that follows. The Toolkit understands the relationship between misogyny and extremism in four interconnected ways:

- **As a form of extremism in its own right:** misogyny can constitute an extremist ideology, often referred to as male supremacism. Analysis of incel-linked and other misogynistic attacks shows that some such violence meets core criteria for violent extremism, including clear ideological motivation and the use of violence to intimidate or coerce wider audiences.
- **As a core component underpinning other forms of extremism:** violent misogyny is present across far-right, far-left and jihadist movements, where it reinforces hierarchies, legitimizes violence and facilitates recruitment. Misogynistic online spaces can also act as gateways into broader extremist networks.
- **As a risk factor:** survey-based studies show that misogynistic attitudes are associated with increased support for violent extremist attitudes, greater willingness to engage in interpersonal violence, and higher tolerance for violence against women.
- **As a warning sign:** many terrorist perpetrators have prior histories of domestic violence or other expressed misogyny. Subsequent attacks are better understood as a continuation of violence than as a shift from 'private' to 'public' violence.

Who this report is for

This report is intended for stakeholders responsible for developing, implementing, overseeing or informing platform approaches to online harms at the intersection of misogyny, targeted hate and violent extremism including:

- Tech companies, particularly teams responsible for platform policy, product, integrity and safety. This includes Trust & Safety professionals working across TFGBV, hate, harassment, violent extremism, dangerous organizations and violence. They can use it to review how their own policies recognize gender-based harm, and to strengthen the connections between policy areas that are often developed and enforced separately.
- Regulators and policymakers responsible for online safety and platform governance. They can use it as an evidence base for understanding how platforms currently address this intersection of harms, and for shaping expectations of how platforms should assess and respond to them.
- Civil society organizations, researchers and practitioners working across TFGBV, targeted hate and violent extremism. They can use it to understand where protections for women and LGBTQI+ people exist in platform rules and where they fall short, and to inform advocacy, research and engagement with platforms.

How to read this report

CHAPTER 1

Methodology

Sets out what was reviewed, the 12 platforms included in the analysis and the criteria used to assess their policies.

CHAPTER 2

The overall policy landscape for gender-based violence harms

Examines how TFGBV is addressed across public platform policies, where different harms are located, whether their gendered dimensions are recognized, and which harms receive stronger or weaker coverage.

CHAPTER 3

Violent extremism and dangerous organizations, hate and violence policies

Examines what happens when TFGBV intersects with these policy areas, including where harms are located, whether their gendered dimensions are recognized, and how forms of TFGBV are captured.

CHAPTER 4

Cross-cutting themes

Draws together key themes and what the findings mean in practice for platform policy development, enforcement and approaches to interconnected harms.

CHAPTER 5

Recommendations

Provides practical recommendations for strengthening platform responses to TFGBV and its intersections with targeted hate and violent extremism.

CONCLUSION AND ANNEX

Summarizes the key takeaways and provides additional supporting detail for the analysis.

Glossary

This report uses the Project Catalyst definitions for Technology-Facilitated Gender-Based Violence (TFGBV), Targeted Hate & Violence (TH&V) and Violent Extremism (VE), which differ from ISD's definitions.

Anti-feminism · The countermovement opposing emancipatory or feminist ideals that uses misogynist strategies and tactics. Anti-feminists oppose gender equality efforts and the democratic negotiation of gender relations.

Dehumanization · The denial of the full humanity of others. This is often achieved by comparisons to vermin, animals, property or the suggestion that targets are subhuman. Dehumanization justifies the mistreatment of dehumanized groups and can be implicated in mass atrocity as well as targeted violence against these groups.

Feminism · ISD draws upon bell hooks' definition of feminism, which establishes it as a "movement to end sexism, sexist exploitation, and oppression."

Hate Speech · ISD classifies hate speech as any activity or content that seeks to dehumanize, demonize, harass, threaten or incite violence against an individual or community on the basis of their race, ethnicity, sexual orientation, gender, sex, disability, migrant status or religion.

Intimate Image Abuse · Following the UNFPA definition, the use of images to coerce, threaten, harass, objectify or abuse. Deepfakes and non-consensual intimate imagery (NCII) created using AI tools are forms of image-based abuse.

Male Supremacy · 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]."[2] It is linked to hegemonic masculinity, which structures patriarchy and describes the "legitimation of unequal gender relations."

Manosphere · The 'manosphere' is an umbrella term that refers to a number of interconnected misogynistic communities. It encompasses multiple types and severities of misogyny—from broader male supremacist discourse to men's rights activism (MRA) and 'involuntary celibates' (incels).

Misogyny · ISD draws upon the definition of Kate Manne, who argues misogyny operates to uphold a patriarchal social order, policing gender norms to ensure that women and marginalized gender identities conform. Misogyny works to justify violence when there is deviation from these norms. Misogyny includes what might be considered a type of deeply held prejudice toward women and marginalized gender identities.

Online Harassment · Defined by Humane Intelligence as "an umbrella term that encompasses a wide range of behaviors via technology intended to silence or intimidate a target. It often includes repeated hostile messages, direct threats of violence, and offensive or abusive comments (including expression of sexism, racism, xenophobia, homophobia, transphobia or ableist prejudices). It may include coordinated pile-on attacks or sustained harassment campaigns by formal groups or informal groups intended to overwhelm the target."

Patriarchy · Patriarchy refers to a social system in which men hold primary power and authority in various aspects of life, including family, politics, economics and culture. It is characterized by male dominance in decision-making, legal structures and social norms, often reinforced by traditions, religious interpretations and societal expectations.

Targeted Hate and Violence (TH&V) · Following the Project Catalyst definition, targeted hate and violence is any activity that dehumanizes other people by seeking to elevate the position of members in one group while diminishing that of non-members based on a particular protected characteristic of those people; advocates for or uses violence to realize that dehumanization; and tolerates, supports, actively calls for, or directly uses violence against civilians or critical civilian infrastructure.

Technology-Facilitated Gender-Based Violence (TFGBV) · UN Women defines TFGBV as any act that is carried out via information and communication technologies or other digital tools, which results in (or is likely to result in) physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms.

Violent Extremism · Activity that promotes ideological, political, or religious aims while advocating for or using violence to achieve those aims, and that tolerates, supports, actively calls for, or directly uses violence against civilians or critical civilian infrastructure. This definition of violent extremism is layered against ISD's social identity-based definition of extremism, which is understood as the advocacy of political and social change in line with a system of belief that claims the superiority and dominance of one identity-based 'in-group' over an 'out-group.' Extremism advances a dehumanizing 'othering' mindset incompatible with pluralism and universal human rights. It can be pursued through violent or non-violent means.

2 This description draws from the Southern Poverty Law Center's definition of male supremacy, outlined in https://www.splcenter.org/resources/extremist-files/male-supremacy/.

PLATFORM POLICY ANALYSIS

Methodology

In this section

Scope	11
Analysis criteria	12
Limitations	13

Methodology

Scope

The analysis draws on 120 policy documents from 12 platforms, accessed in June 2026. These include Terms of Service, community guidelines and standards, and dedicated policy explainers. A policy document is defined as any separately named policy published by a platform, with jurisdiction-specific versions of Terms of Service counted separately.[3] As platforms organize their rules differently across documents, the number of documents per platform is not used as a measure of policy coverage. The full sample, including hyperlinks to the source documents, is listed in the Annex.

The platforms include TikTok, X, Facebook, Instagram, YouTube, Twitch, Reddit, Discord, Rumble, Telegram, 4chan and Incels.is. These platforms were selected based on the findings of the accompanying Project Catalyst research, which examined the spread of online misogyny and TFGBV across online ecosystems in Canada and Jordan. In addition to each platform's Terms of Service, including separate jurisdiction-specific versions where applicable, ISD selected individual policies and relevant sections addressing violent extremism, targeted hate, violence and TFGBV.

These policies were assessed against the five TFGBV harm categories (or 'super-types') outlined in ISD's Toolkit, Addressing the Intersection of Misogyny, Targeted Hate and Violent Extremism: online harassment, intimate image abuse, online impersonation, sexual extortion and account access control. These are summarized in Box 1 below. Assessing platform policy against these categories makes it possible to identify which harms are explicitly recognized as gendered, which are addressed only in gender-neutral terms, and which are not covered by any platform policy. The analysis also serves as a baseline for the Toolkit's recommendations to tech platforms by establishing current policy language around TFGBV, violent extremism, targeted hate, as well as violence more generally.

BOX 1 The Five TFGBV 'Super-Types'

SUPER-TYPE	COMMON TACTICS AND SUB-HARMS	ILLUSTRATIVE EXAMPLES OF INTERSECTIONS WITH VIOLENT EXTREMISM (PER THE TOOLKIT)
Online Harassment	Doxxing, cyberstalking, dogpiling or mobbing, gendered hate speech, coordinated and cross-platform harassment	Online harassment can intersect with violent extremism through tactics such as hate speech, incitement and threats of violence.
Intimate Image Abuse	Non-consensual intimate imagery, sexual deepfakes, voyeuristic recording, cyberflashing	This can involve coercion linked to supremacist beliefs or the funding of extremist activities through exploitation.
Online Impersonation	Catfishing, fake profiles, account cloning, deepfake impersonation	This can include impersonation to advance supremacist beliefs or violent extremist agendas including human trafficking or terrorist recruitment.

3 Documents were coded at the level of the named policy or section, resulting in 121 coded entries. TikTok's US Terms of Service was coded as two separate sections.

SUPER-TYPE	COMMON TACTICS AND SUB-HARMS	ILLUSTRATIVE EXAMPLES OF INTERSECTIONS WITH VIOLENT EXTREMISM (PER THE TOOLKIT)
Sexual Extortion	Sexual blackmail, online sexual coercion, grooming, live-streamed sexual exploitation	The use of sexual extortion as a coercive tactic to advance extremist ideology, including demanding sexual acts, images, or compliance under threat of exposure, where such threats are explicitly linked to male-supremacist or far-right narratives that justify domination or violence.
Account Access Control	Hacking, account lockout, spyware and tracking, outing of gender identity or sexual orientation	This can include the exploitation of system or account access to facilitate physical harm or operational support, including the disclosure of real-world locations, security vulnerabilities, or sensitive personal or organizational information, including women's health care and abortion providers, fertility clinics, and gender affirming care for trans communities, thereby enabling or accelerating ideologically motivated violence.

Source: ISD, Addressing the Intersection of Misogyny, Targeted Hate and Violent Extremism (Toolkit). pp. 12–16.

Analysis criteria

The analysis was conducted in three stages. First, all relevant policies published by each platform were collected and assessed against the five TFGBV harm categories (super-types). This included Terms of Service, community guidelines and individual policy explainers.[4] Each policy was recorded with its coverage of each harm, how it framed the harm, any gender-related gaps and whether it carried a violent extremism or dangerous organizations dimension. Assessing policies individually, rather than at the platform level, follows the Toolkit's guidance to examine these harms across all relevant policy areas. These may include hate speech, violent extremism and dangerous organizations, threats and violence, harassment and bullying, and sexual abuse, rather than assuming that a particular harm falls within a single policy area.

Second, each platform was assessed based on how its policies address each type of gender-based harm. A harm (e.g. online harassment) is considered fully covered where it is addressed as a behavior, tactic or sub-harm (e.g. doxxing) and its gender dimension is recognized, either explicitly through protected-characteristic framing or because the harm is inherently sexual or consent-based (e.g. sexual violence). Coverage is partial where the harm, tactic or behavior is addressed but only in gender-neutral terms, and absent where it is not represented in any policy.

This mirrors two of the guiding questions the Toolkit asks platforms and industry stakeholders to consider when assessing the intersection of TFGBV, targeted hate and violent extremism: whether the behavior falls within one of the five TFGBV harm categories, and whether it is sexualized or gender related. A policy that answers the first question but not the second recognizes the conduct itself without acknowledging its gendered dimension. This is the distinction captured by a partial coverage rating.

Accordingly, this analysis treats gender-neutral coverage of a TFGBV-related harm, such as online harassment, as incomplete ('partial') rather than sufficient. The Toolkit's comparative legal analysis[5] identifies account access control and online impersonation as harms commonly treated as gender-neutral technical offenses despite their gendered use and impact. It also highlights how the framework under which a harm is categorized shapes

4 While Terms of Service set out the agreement between platform and user, Community Guidelines serve as the user-facing expression of a platform's content and conduct rules, written to set expectations and educate users. Policy explainers elaborate a single policy area in greater detail, typically including definitions, examples and scope conditions.

how it is understood, resourced and addressed. The same applies to platform policies: where a harm is addressed determines which policy governs it and, in turn, what action a platform can take in response.

A harm addressed only in gender-neutral terms diverges from the Toolkit's recommendations on recognizing the specific characteristics of gender-based violence. These include recognizing male-supremacist extremism and gendered hate mobilization within platform policy frameworks, identifying the specific targeting of women and LGBTQI+ people, and ensuring case triage across harassment, hate and extremism teams. A high-level, distilled summary of these findings is presented in Table 1 below.

The subset of policies focused on violent extremism and dangerous organizations; hate and violence was then examined individually. The Toolkit recommends that platforms recognize male-supremacist extremism and gendered hate mobilization as enforceable categories within their existing policy frameworks. This analysis examines publicly available platform policies to establish the extent to which these categories are currently recognized. Three things were recorded for each policy: which TFGBV harms it fully captures, whether it references misogyny and how it engages the gender dimension. Policies were found to address the gender dimension in one of three ways. Some name it directly, listing gender among their protected characteristics. Some borrow it, invoking protected categories that are defined only in a separate policy. Others capture an inherently sexual or consent-based harm without naming protected categories at all. Table 2 presents high-level takeaways from this analysis.

The annex lists all platform policies assessed.

Limitations

This analysis considers only publicly available policies published by each platform. Published policy language differs from the internal guidance that enforcement teams apply, and the analysis does not assess how policies are enforced in practice. All policies were accessed in June 2026. Where platforms publish separate versions of their Terms of Service by jurisdiction, each version was reviewed and counted separately; community guidelines and policy explainers were treated as applying globally. English-language versions of all policies were reviewed.

5 A systematic review was conducted of relevant regulatory frameworks in Canada, Kenya, and Jordan (the focus jurisdictions of Project Catalyst), to assess how policy and regulation apply across the harms identified in the taxonomy. This analysis identifies and assesses legal levers available when responding to different parts of the continuum of gender-based harm. This legal analysis translates the harms taxonomy into practical guidance on applying existing laws and policy frameworks to misogyny across a continuum of violence from TFGBV to targeted hate and violent extremism. It also outlines how different legal systems recognize, categorize and respond to these harms, as well as where approaches might be falling short.

The Overall Policy Landscape for Gender-Based Violence Harms

In this section

Table 1: TFGBV Harm Coverage Across 12 Platforms	**15**
Where harms are located	**17**
Gender dimension	**17**
Which harms are better covered	**18**

The two sections below use the same three guiding questions: (1) where harms are located across platform policies, (2) how their gendered dimensions are treated and (3) how comprehensively they are addressed. The analysis first examines the broader policy landscape before focusing specifically on policies addressing violent extremism and dangerous organizations, hate and violence.

The Overall Policy Landscape for Gender-Based Violence Harms

Table 1 maps, for each of the 12 platforms, whether each of the five TFGBV harms is covered with its gender dimension recognized ('Covered'), addressed only in gender-neutral terms ('Partial') or not addressed in any policy ('Not covered'). Each cell shows the highest rating a platform achieves for a given harm across all its policies and identifies the policies receiving that rating. A Covered cell lists the policies that fully capture the harm, while a Partial cell lists those that address it in gender-neutral terms. Where a harm is fully captured by at least one policy, other policies may also address it in gender-neutral terms (and therefore receive a partial rating). However, these policies are not listed in the table.

TABLE 1

TFGBV Harm Coverage Across 12 Platforms

TABLE GUIDANCE

Covered harm addressed and its gender-based character recognized

Partial harm addressed, but in gender-neutral terms

Not covered not addressed in any policy

Platform	Online Harassment	Intimate Image Abuse	Online Impersonation	Sexual Extortion	Account Access Control
TikTok	Covered *Community Guidelines (Hate Speech and Hateful Behavior; Harassment and Bullying); Terms of Service [European Economic Area (EEA), United Kingdom (UK), or Switzerland (CH); Other Regions outside of the United States]*	Covered *Community Guidelines [Adult Sexual Abuse; Integrity and Authenticity - Edited Media and AI-Generated Content (AIGC); Accounts and Features]*	Partial *Terms of Service (EEA/UK/CH); Terms of Service (US); Terms of Service (If you are not in the US, EEA, the United Kingdom, Switzerland or India); Community Guidelines [Integrity and Authenticity - Edited Media and AI-Generated Content (AIGC); Integrity and Authenticity—Deceptive Behaviors and Fake Engagement; Privacy and Security - Personal Information]*	Covered *Community Guidelines (Youth Sexual and Physical Abuse; Adult Sexual Abuse; Human Trafficking and Smuggling; Harassment and Bullying; Accounts and Features)*	Partial *Terms of Service (EEA/UK/CH); Terms of Service (US); Terms of Service (If you are not in the US, EEA, the United Kingdom, Switzerland or India); Community Guidelines (Privacy and Security - Platform Security)*
X	Covered *Rules and Policies (Violent and hateful entities policy; Hateful Conduct; Abuse and Harassment; Violent Content; Abuse profile information)*	Covered *Rules and Policies (Private Content; Violent Content; Non-consensual nudity policy; Child Safety)*	Partial *Terms of Service (EU, EFTA States, or UK); Terms of Service - US, and other jurisdictions; Rules and Policies - Authenticity*	Covered *Rules and Policies (Child Safety)*	Partial *Rules and Policies (Private Content; Authenticity)*
Facebook	Covered *Meta Community Standards (Dangerous organizations and individuals; Violence and incitement; Adult sexual exploitation; Bullying and harassment; Child sexual exploitation, abuse and nudity; Hateful Conduct)*	Covered *Meta Community Standards (Adult sexual exploitation; Child sexual exploitation, abuse and nudity)*	Partial *Meta Community Standards (Coordinating harm and promoting crime; Authentic identity representation; Inauthentic Behavior; Fraud, scams and deceptive practices)*	Covered *Meta Community Standards (Adult sexual exploitation; Child sexual exploitation, abuse and nudity; Human Exploitation)*	Partial *Terms of Service; Meta Community Standards (Privacy Violations; Cybersecurity)*

Platform	Online Harassment	Intimate Image Abuse	Online Impersonation	Sexual Extortion	Account Access Control
Instagram	Covered *Meta Community Standards (Dangerous organizations and individuals; Violence and incitement; Adult sexual exploitation; Bullying and harassment; Child sexual exploitation, abuse and nudity; Hateful Conduct)*	Covered *Meta Community Standards (Adult sexual exploitation; Child sexual exploitation, abuse and nudity)*	Partial *Instagram Terms of Use; Meta Community Standards (Coordinating harm and promoting crime; Authentic identity representation; Inauthentic Behavior; Fraud, scams and deceptive practices)*	Covered *Meta Community Standards (Adult sexual exploitation; Child sexual exploitation, abuse and nudity; Human Exploitation)*	Partial *Instagram Terms of Use; Meta Community Standards (Privacy Violations; Cybersecurity)*
YouTube	Covered *Hate speech policy; Harassment & Cyberbullying policies*	Covered *Nudity & Sexual Content Policy*	Partial *Policies - Spam & deceptive practices - Impersonation policy*	Partial *Policies - Sensitive Content - Child safety policy*	Partial *Violent or dangerous content - Harmful or dangerous content policy*
Twitch	Covered *Community Guidelines (Violence and Threats; Adult Sexual Violence; Hateful Conduct; Sexual Harassment; Account Usernames and Display Names; Harmful Misinformation Actors)*	Covered *Community Guidelines (Youth Safety; Off-Service Conduct; Sexual Harassment)*	Partial *Terms of Service; Community Guidelines - Account Usernames and Display Names; Impersonation*	Covered *Community Guidelines (Adult Sexual Violence; Youth Safety; Off-Service Conduct; Sexual Harassment)*	Partial *Terms of Service; Community Guidelines - Violence and Threats*
Reddit	Covered *Harassment; Hate policies*	Covered *Reddit Rules; Harassment; Never Post Intimate or Sexually Explicit Media of Someone Without Their Consent*	Partial *Reddit Rules; Manipulated Content and Misleading Behavior*	Partial *Reddit Rules; Illegal or Prohibited Transactions*	Partial *User Agreement - Reddit User Agreement if you live in the United States or outside the EEA, United Kingdom, and Switzerland; User Agreement - Reddit User Agreement if you live in the EEA, United Kingdom, or Switzerland; Illegal or Prohibited Transactions*
Discord	Covered *Community Guidelines; Bullying, Harassment, and Threats Policy Explainer; Hateful Conduct Policy Explainer*	Covered *Community Guidelines; Bullying, Harassment, and Threats Policy Explainer; Non-Consensual Intimate Imagery Policy Explainer; Teen and Child Safety Policy Explainer; Discord Developer Policy*	Partial *Community Guidelines; Identity and Authenticity Policy Explainer; Deceptive Practices Policy Explainer*	Covered *Bullying, Harassment, and Threats Policy Explainer; Sexual Content Policy Explainer; Non-Consensual Intimate Imagery Policy Explainer; Teen and Child Safety Policy Explainer; Human Trafficking Policy Explainer*	Partial *Terms of Service; Bullying, Harassment, and Threats Policy Explainer*
Rumble	Partial *Website Terms and Conditions of Use and Agency Agreement*	Partial *Website Terms and Conditions of Use and Agency Agreement*	Not covered	Partial *Website Terms and Conditions of Use and Agency Agreement*	Not covered
Telegram	Partial *User guidance for the EU Digital Services Act*	Covered *User guidance for the EU Digital Services Act*	Not covered	Partial *Terms of Service*	Not covered
4chan	Partial *Rules - Global Rules*	Not covered	Not covered	Not covered	Not covered
Incels.is	Partial *Terms and Rules*	Not covered	Not covered	Not covered	Not covered

Where harms are located

Eight of the 12 platforms analyzed (TikTok, X, Facebook, Instagram, YouTube, Twitch, Reddit and Discord) have dedicated policies addressing violent extremism and dangerous organizations, hate and violence, alongside other policies specifically addressing TFGBV-related harms. Four platforms, namely Rumble, Telegram, 4chan and Incels.is, only maintain general rules or Terms of Service, though Telegram additionally includes user guidance under the EU Digital Services Act.[6]

Across TikTok, Facebook, X, Instagram, YouTube, Twitch, Reddit and Discord, TFGBV protections are largely found in community guidelines and policy explainers rather than in Terms of Service. This is unsurprising, given that Terms of Service generally set out broader rules governing platform use. TikTok's Terms of Service are the only ones identified in this analysis that fully capture a TFGBV harm at the terms level, prohibiting harassment and discrimination on the basis of sex, gender identity and sexual orientation in its EEA, UK and Switzerland, and rest-of-world terms. However, its US Terms of Service contain no harassment prohibition, demonstrating how protections at the terms level can vary by jurisdiction.

Coverage is scattered given a single harm can be found spread across many policies. Sexual extortion on TikTok, for example, is included in five separate policies within the community guidelines (in Youth Sexual and Physical Abuse; Adult Sexual Abuse; Human Trafficking and Smuggling; Harassment and Bullying; Accounts and Features).

Three of the 12 platforms extend enforcement to off-platform conduct, with different TFGBV implications and reach. Twitch's Off-Service Conduct names sexual assault, grooming and non-consensual sexual activity, TikTok bans accounts for off-platform "act[s] of sexual abuse against a youth or adult" alongside violent or hateful-organization membership, while Discord's Off-Platform Behaviors Policy Explainer is focused on physical harm, with no sexual or gendered element, though it is invoked by reference in Discord's child safety and violent extremism policies.[7]

Gender dimension

Policies analyzed across platforms largely capture the behaviors of TFGBV but lack the gender dimension. Harassment (and the sub-harm of doxxing in particular), impersonation and account compromise are widely addressed, but predominantly in gender-neutral terms. Where gender is recognized, it is primarily framed either as a protected characteristic (or 'attribute,' 'group,' or 'target'), or through the inherently sexual or consent-based nature of the harm, as in cases involving non-consensual intimate imagery, sexual extortion, threats of sexual assault and depictions of sexual violence. Policies prohibiting these harms capture gender-based violence without naming gender, because the gender dimension is carried by the nature of the conduct rather than by the identity of the target. Rumble, Telegram, Incels.is and 4chan do not include protected characteristics in their rules.

6 User guidance for the EU Digital Services Act.
7 Discord's Off-Platform Behaviors Policy Explainer itself defines its scope in exclusively physical-harm terms: "This policy covers activity involving the physical harm of others or that presents a risk of causing physical harm to others." However, Discord's separate Teen and Child Safety policy and Violent Extremism policy each state that they will draw on this Off-Platform Behaviors framework when assessing off-platform conduct in their own domains. The Child Safety Policy noting that "we will also consider off-platform evidence as explained in our Off-Platform Behaviors Policy" for child sexualization and grooming content "due to the high-harm nature of the offense," and the Violent Extremism stating that "violent extremist organizing is also subject to our Off-Platform Behaviors Policy." Therefore, the Off-Platform Behaviors Policy's own stated scope contains no sexual or gendered element, but it is incorporated by reference into at least two other Discord policies that do address sexual harm (against minors) and extremism.

Which harms are better covered

Harassment is the only harm that every platform includes to some extent. TikTok, X, Facebook, Instagram, YouTube, Twitch, Reddit and Discord include harassment fully through a gender dimension. Telegram, Rumble, 4chan and Incels.is only include harassment partially, lacking the gender dimension. Their policies provide little further coverage of TFGBV-related harms.

Intimate image abuse is the next most comprehensively covered harm. Nine of the 12 platforms (TikTok, X, Facebook, Instagram, YouTube, Twitch, Reddit, Discord, and Telegram) fully capture this harm and its gendered dimensions. For Telegram, this coverage is found in its EU Digital Services Act user guidance. Some platforms also have dedicated policies addressing intimate image abuse, including X's Non-Consensual Nudity policy, Discord's Non-Consensual Intimate Imagery explainer, and Reddit's standalone rule, "Never Post Intimate or Sexually Explicit Media of Someone Without Their Consent."

Sexual extortion is fully captured on six platforms (TikTok, X, Facebook, Instagram, Twitch and Discord), but for some platforms' coverage, it is mainly related to youth policies**,** therefore at times lacking protection of sexual extortion for adults. For example, X's only coverage of sexual extortion is located in its Child Safety policy and YouTube's sole (partial) coverage is also child safety focused. However, TikTok, Twitch, Instagram, Facebook and Discord cover the harm for both adults and minors.

Impersonation and account access control are never fully captured with their gendered dimension recognized in the policies of any platform included in the analysis. Both harms are partial on TikTok, YouTube, Reddit, X, Facebook, Instagram, Twitch and Discord, and absent on Rumble, Telegram, 4chan and Incels.is. Despite being established TFGBV tactics, both harms are largely handled through spam, authenticity and security provisions, where protected-characteristic framing is not used. Their gendered dimensions are therefore recognized only where a platform explicitly names them, which none of the platforms in this analysis does. Account access control in particular is frequently located in Terms of Service (as on TikTok and Reddit), whereas full captures of the other harms mostly sit in Community Guidelines. This mirrors a pattern identified in the Toolkit's legal analysis of the regulatory landscapes in Canada, Jordan and Kenya: account access control and online impersonation are commonly treated as gender-neutral technical offenses, despite their gendered use and impact.

Violent Extremism and Dangerous Organizations, Hate and Violence Policies: Coverage of TFGBV Harms

In this section

Table 2: Violent Extremism and Dangerous Organizations, Hate, & Violence Policies	20
Where harms are located	23
Gender dimension	23
Which harms are captured and how	24
Additional Policies	24

This section examines where TFGBV harms are addressed across each platform's rules, focusing on a subset of policies governing violent extremism and dangerous organizations, hate and violence. These policy areas are particularly relevant to the enforcement of rules addressing the intersections between TFGBV, targeted hate and violent extremism. They are also the frameworks the Toolkit asks platforms to use when recognizing male-supremacist extremism and gendered hate mobilization, and when developing or adapting policies to reflect the varied ways in which misogyny and TFGBV manifest. This includes ensuring that harms are assessed across all relevant policy areas, rather than treated in isolation.

Violent Extremism and Dangerous Organizations, Hate and Violence Policies

A policy can engage with the gendered dimensions of harm in several ways. It may recognize gender as a basis on which people are targeted by including gender, gender identity, sexual orientation, gender expression, sex characteristics or non-adherence to prescribed gender roles or norms, as outlined in the Toolkit, among its protected characteristics. Alternatively, a policy may explicitly prohibit gender-based harms, identify misogyny as a hateful ideology or address conduct that is inherently sexual or consent-based, such as non-consensual sexual imagery or threats of sexual assault. In these cases, the gendered nature of the harm is recognized through the conduct itself, rather than the identity of the target.

Table 2 shows which TFGBV harms are fully covered by the violent extremism and dangerous organizations, hate and violence policies, and how their gendered dimensions are recognized. A harm is considered fully covered where a policy both prohibits the conduct and recognizes its gendered dimension, either by naming gender among the characteristics it protects or because the prohibited conduct is inherently sexual or consent based. Each entry identifies the harm, the policy that covers it, how the policy recognizes its gendered dimension and an example of the relevant policy language. Where these policies prohibit a harm without recognizing its gendered dimension, coverage is considered partial and is therefore not shown in this table.

For the purposes of this analysis, these policies were grouped into three categories by what each primarily does: policies that include violent extremism or related dangerous organizations/individuals; those built around hateful conduct, hate speech and protected characteristics as hate policies; and those prohibiting violent acts or content as violence policies. The first column lists the protected characteristics each platform names that relate to gender. Platforms protect a wider range of characteristics; only the gender-related ones are shown, since these determine what the policies in this table can recognize as gender-based harm.

TABLE 2

Violent Extremism and Dangerous Organizations, Hate, & Violence Policies

TABLE GUIDANCE

Gender named the policy names the gender dimension itself

Sexual/consent-based harm inherently sexual/consent-based harm captured

None fully policy exists, no TFGBV-harm fully captured

Gender borrowed protected categories including gender used, but defined in another policy

N/A no relevant policy

Platform (and protected characteristics if present)	POLICY TYPE		
	Violent Extremism / Dangerous Organizations	**Hate**	**Violence**
TikTok *(Protected attributes include gender, gender identity, sex, sexual orientation.)*	**Online Harassment** Gender borrowed *"Hateful Organizations: Groups that target people based on protected attributes, dehumanize others, and promote hateful ideologies."* ***Violent and Hateful Organizations and Individuals***	**Online Harassment** Gender named *"Encouraging violence, segregation, discrimination, or other harms towards people based on protected attributes."* Misogyny named within list of hateful ideologies: *"Claims of supremacy over a protected group, such as white supremacy, misogyny, anti-LGBTQ+ hate."* ***Hate Speech and Hateful Behavior***	None fully *Threats & incitement without gender dimension.* ***Violent and Criminal Behavior***

Platform (and protected characteristics if present)	POLICY TYPE **Violent Extremism / Dangerous Organizations**	**Hate**	**Violence**
X *(Protected categories include gender, gender identity, sexual orientation.)*	**Online Harassment** Gender borrowed *"Hateful entities are those that have systematically and intentionally promoted, supported and/or advocated for hateful conduct, which includes promoting violence or engaging in targeted harassment towards a protected category."* ***Violent and Hateful Entities***	**Online Harassment** Gender named *"Targeting others with repeated slurs, tropes or other content that intends to degrade or reinforce negative or harmful stereotypes about a protected category... racist/sexist tropes where the context is to harass or intimidate others"* *"Incitement: inciting others to harass members of a protected category on or off platform"* *"Slurs and Tropes: racist/sexist tropes where the context is to harass or intimidate."* *"Dehumanization: dehumanization of a group of people based on... gender, gender identity, or sexual orientation"* *"Hateful Profile: abusive behavior targeting groups or protected categories."* ***Hateful Conduct***	**Online Harassment** Sexual/consent-based harm Threats to sexually assault *"Violent Threats: Threats to inflict physical harm on others, which includes threatening to kill, torture, sexually assault, or otherwise hurt someone."* ***Violent Content*** **Intimate Image Abuse** Sexual/consent-based harm Media of non-consensual sexual acts: *"Violent Sexual Conduct: media that depicts violence, whether real or simulated, in association with sexual acts, including rape and other forms of violent sexual assault such as sexual acts that occur without the consent of participants, and sexualized violence within an intimate setting, where it is not immediately obvious if those involved have consented to take part."* ***Violent Content***
Facebook & Instagram (Meta) *(Protected characteristics include sexual orientation, sex, gender identity.)*	**Online Harassment** Gender borrowed 'Hate entity' definition includes *"Violence, threatening rhetoric or dangerous forms of harassment targeting people based on their protected characteristics"* ***Dangerous organizations and individuals***	**Online Harassment** Gender named Hateful conduct defined as: *"direct attacks against people... on the basis of what we call protected characteristics... sexual orientation, sex, gender identity"* *"We remove dehumanizing speech, allegations of serious immorality or criminality, and slurs. We also remove harmful stereotypes, which we define as dehumanizing comparisons that have historically been used to attack, intimidate or exclude specific groups, and that are often linked with offline violence. Finally, we remove serious insults, expressions of contempt or disgust, swearing and calls for exclusion or segregation when targeting people based on protected characteristics."* ***Hateful Conduct***	**Online Harassment** Gender named *"Glorification of gender-based violence that is either intimate partner violence or honor-based violence."* *"Violent speech targeting a person or group of people on the basis of their protected characteristic(s)."* ***Violence and Incitement***
YouTube *(Protected group status includes sex, gender, sexual orientation.)*	None fully *Incitement against "a defined group of people" - no protected groups named.* ***Violent Extremist or Criminal Organizations Policy***	**Online Harassment** Gender named *"Repeatedly targets, insults, and abuses a group based on protected group status across multiple uploads."* *"Exposes a group with protected group status to risks of physical harm based on the local social or political context."* *"Encourage violence against individuals or groups based on their protected group status."* ***Hate Speech Policy***	**Intimate Image Abuse** Sexual/consent-based harm *"Violent physical sexual assaults (video, still imagery, or audio)' not allowed."* ***Violent or Graphic Content Policies***

Platform (and protected characteristics if present)	POLICY TYPE **Violent Extremism / Dangerous Organizations**	**Hate**	**Violence**
Twitch *(Protected characteristics include sex, gender, gender identity, sexual orientation.)*	None fully *Content that depicts, glorifies, encourages, or supports terrorism, or violent extremist actors or acts; threatening to or encouraging others to commit acts that would result in serious physical harm or significant property destruction. This does not include mention of protected groups.* ***Terrorism and Violent Extremism***	**Online Harassment** Gender named *"Promote, glorify, threaten, or advocate violence, physical harm, or death against individual(s) or groups on the basis of a protected characteristic"* *"Intentionally referring to someone using a pronoun or form of address that does not correctly reflect the gender with which they identify, such as repeating incorrect pronouns after being asked to stop."* *"Calls to increase fertility rates or restrict immigration to preserve the economic/political/social power of any group."* *"Incite or promote violence against a protected group."* ***Hateful Conduct***	**Online Harassment** Sexual/consent-based harm *"Content or activities that threaten or promote sexual violence are strictly prohibited and may be reported to law enforcement."* ***Violence and Threats*** **Sexual Extortion** Sexual/consent-based harm *"Coercing a guest into committing a sex act by threatening them"* ***Adult Sexual Violence***[8]
Reddit *(Protected characteristics include gender, gender identity, sex, sexual orientation, victims of violent crimes and their families, pregnancy.)*	None fully *Recruitment and incitement without protected groups.* ***How does Reddit fight the dissemination of terrorist content?***	**Online Harassment** Gender named *"Engaging in a pattern of abuse, site-stalking, or coordinating attacks to intimidate an individual or antagonize a community of people, on the basis of their identity or vulnerability."* *"A comment arguing that raping women should be acceptable and not a crime."* *"Directly replying to someone's post to attack their gender."* *"A comment stating that a specific individual is too stupid to lead a country because of their gender or racial identity."* ***Hate Policy***	**Intimate Image Abuse** Sexual/consent-based harm *"Promoting real-world sexual violence or sexual activity that involves severe physical injury or risk of death."* ***Violence Policy***
Discord *(Protected characteristics include gender, gender identity, sex, sexual orientation, status as a victim of domestic/ sexual violence or stalking.)*	None fully *Names harassment and doxxing within VE tactics without gender dimension.* ***Violent Extremism Policy Explainer***	**Online Harassment** Gender named *"We define hate speech as any expression that degrades, vilifies, or dehumanizes individuals, incites hostility towards specific groups, or promotes harm based on protected characteristics."* ***Hateful Conduct***	**Intimate Image Abuse** Sexual/consent-based harm *"Promoting, celebrating, or inciting sexual violence against others."* ***Violence and Graphic Content Policy Explainer***
Rumble	N/A[9]	N/A	N/A
Telegram	N/A[10]	N/A	N/A
4chan	N/A	N/A	N/A
Incels.is	N/A	N/A	N/A

8 Twitch's Off-Service Conduct policy also names sexual assault and grooming but is a scope policy (defining off-platform enforcement) rather than a violence policy; it is listed in the annex, not counted in the violence family here.
9 Rumble's Terms and Conditions do contain an entity-designation clause, prohibiting content that "promotes, supports or incites individuals and/or groups which engage in violence or unlawful acts, including but not limited to Antifa... the KKK and white supremacist groups," and content supporting entities designated as terrorist by the Canadian or US governments. It is recorded as N/A because this is a clause within a general terms-of-service document rather than a standalone policy, with no definitions, behavioral provisions or enforcement mechanism attached.
10 Telegram's Terms of Service prohibit "promot[ing] violence on publicly viewable Telegram channels," and its EU DSA guidance adds a prohibition on "calls for violence, terrorism, soliciting funds for terrorist organizations." Both are recorded as N/A as they are general clauses rather than dedicated policies, with no definitions, designation framework or stated enforcement mechanism.

Where harms are located

Within the policies relating to violent extremism and dangerous organizations, hate and violence, **TFGBV protection concentrates in hate policies as well as some violence policies, while being less present in violent extremism and dangerous organizations policies.** The hate policy column in Table 2 demonstrates how hate policies on eight platforms (TikTok, X, YouTube, Instagram, Facebook, Twitch, Reddit and Discord) all fully capture online harassment with the gender dimension named directly. Only three violent extremism and dangerous organizations policies reflect TFGBV harms, including TikTok, Meta (Instagram and Facebook) and X. For each of these policies, entity definitions and recognition of the gender dimension rely on a 'protected category' that includes gender, but is defined and named only in the platform's separate hate policy.

The remaining dedicated violent extremism and dangerous organizations policies address conduct that overlaps with TFGBV tactics, but not who those tactics are used against. None connects them to gender-based harm or misogyny, and no male-supremacist entity is named in any published policy or entity list. The Toolkit identifies four relationships between misogyny and extremism—misogyny as a form of extremism in its own right, as a core component of other extremist movements, as a risk factor, and as a warning sign of progression toward violence—and none is reflected in how these remaining policies define their own scope.

Within that scope, TFGBV-relevant tactics are addressed only in gender-neutral terms. YouTube's violent extremist or criminal organizations policy prohibits inciting violence against a "defined group of people" but does not name protected groups, relying instead on government and international organization designations; Twitch's Terrorism and Violent Extremism policy includes content that depicts, glorifies, encourages, or supports terrorism, or violent extremist actors or acts and threatening to or encouraging others to commit acts that would result in serious physical harm or significant property destruction without reference to protected characteristics; Discord's Violent Extremism Policy Explainer names harassment and doxxing as extremist tactics but with no gender basis; and Reddit's terrorist content guidance covers glorification, incitement and recruitment but does not name protected characteristics.

The omission of a gender dimension matters when the policy does not address who these tactics are used against. A coordinated campaign of harassment and doxxing conducted by a male-supremacist network would, absent that link, likely be assessed as a series of individual harassment cases under a separate policy.

Among the platforms without dedicated policies, 4chan and Incels.is offer only general rules, Telegram's DSA guidance adds a prohibition on "calls for violence, terrorism, soliciting funds for terrorist organizations," and Rumble names a select few specific entities (including Antifa, the KKK, white supremacist groups and government-designated terrorists).

Gender dimension

The three policy categories address the gender dimension very differently. **Hate policies largely recognize gender directly**, via protected characteristics (or 'groups,' 'attributes' and 'targets') naming gender, gender identity, sex and sexual orientation. Violence policies typically capture TFGBV harms through more inherently sexual, consent-based harms such as naming sexual violence and non-consensual acts within their policies. **Violent extremism and dangerous organizations policies rarely name the gender dimension directly**. Instead, the policies refer to a 'protected category,' which is then defined separately in a hate policy to include a gender dimension.

Two policies provide unique examples of platforms naming a gendered harm or ideology in its own terms rather than reaching it through a protected characteristics list or through the nature of the harm. This includes Meta's Violence and Incitement policy, which is the sole violence policy analyzed to prohibit gender-based violence as such, which includes "glorification of gender-based violence that is either intimate partner violence or honor-based violence." TikTok's Hate Speech and Hateful Behavior policy names misogyny as an ideology listed among "claims of supremacy over a protected group, such as white supremacy, misogyny, anti-LGBTQ+ hate." This contrasts Meta's Dangerous Organizations and Individuals policy, which recognizes "certain ideologies and beliefs that are inherently tied to violence and attempts to organize people around calls for violence or exclusion of others based on their protected characteristics." It names Nazism, white supremacy, white separatism and white nationalism, with misogyny absent from that list.

Which harms are captured and how

Only three of the five TFGBV harm types are fully covered by these policies, with their gender-based character recognized. Online harassment is the most frequently included, reflected in all eight platforms that maintain these policies; primarily through hate policies, occasionally also through violence policies' sexual-threat language (X, Twitch). Intimate image abuse is fully reflected only through violence policies (X, YouTube, Reddit, Discord), via their policies including depiction of sexual violence. Sexual extortion appears once (Twitch's Adult Sexual Violence), while impersonation and account access control are never fully captured.

Within violent extremism and dangerous organizations policies in particular, harassment is the only TFGBV harm that appears. Across all violent extremism and dangerous organizations policies, intimate image abuse, impersonation and account access control are absent and sexual extortion only surfaces once (as partial, with no gender dimension), through TikTok's Violent Criminal Organizations clause covering trafficking. The Toolkit, by contrast, sets out violent extremism indicators and examples for all five harm types. For example, it identifies sexual extortion as a coercive tactic advancing extremist ideology, citing 764, whose members used coerced child sexual abuse material to extort primarily minor victims and to force acts of violence or self-harm; and impersonation as a recruitment and radicalization tactic, citing ISIS recruiters who impersonated women to recruit young Western women, who were then forced into marriage and sexual exploitation.

Additional Policies

Two policies are relevant but fit none of the above categorization for violent extremism, hate or violence but are worth noting for what they capture. This includes Meta's Coordinating Harm and Twitch's Off-Service Conduct, shown below.

TWITCH

Policy: Off-Service Conduct

"We do not allow individuals or organizations who engage off-service in: ... Carrying out or deliberately acting as an accomplice to non-consensual sexual activities and/or sexual assault; Sexual exploitation of youth, such as grooming and/or solicitation/ distribution of CSAM", listed alongside "deadly violence and violent extremism," "terrorist activities or recruiting," "explicit and/or credible threats of mass violence (i.e. threats against a group of people, event, or location where people would gather)," as well as "leadership, membership, or sponsorship of a known hate group."

FACEBOOK & INSTAGRAM (META)

Policy: Coordinating Harm & Promoting Crime

"For the following Community Standards, we require additional information and/or context to enforce: We remove: Outing: exposing the identity of a person and putting them at risk of harm: LGBTQIA+ members; Unveiled women..."

Meta's Coordinating Harm and Promoting Crime policy prohibits "outing," defined as "exposing the identity or locations affiliated with anyone who is alleged to be a member of an outing-risk group." These groups include LGBTQIA+ people and "unveiled women." This recognizes that, for some groups, exposure itself constitutes harm because identification carries a risk of violence. Twitch's Off-Service Conduct policy similarly places sexual harms and extremist affiliation within the same grouping of disqualifying conduct. Carrying out or acting as an accomplice to non-consensual sexual activities or sexual assault, and sexual exploitation of youth such as grooming, are listed alongside deadly violence, violent extremism, terrorist activities and membership of a known hate group. This treats gendered sexual harm and extremist violence under the same umbrella of risk and closely relates to the continuum of harm the Toolkit describes.

④

Cross-cutting themes

Taken together, the findings point to several themes that cut across platforms and policy areas. These inform the recommendations set out in the next section.

Firstly, violent extremism and dangerous organizations policies rarely define gender themselves. Where they protect people from being targeted on the basis of a characteristic, they generally do so by invoking the platform's list of protected characteristics, which is outlined or defined in a separate hate policy. Therefore, these policies recognize gender only to the extent that the hate policy does. No platform's protected characteristics include gender expression, sex characteristics or non-adherence to prescribed gender roles or norms. Trans and non-binary people may be targeted because their identities or expression are perceived to challenge prevailing gender norms and hierarchies. However, currently such targeting is only captured where it can be attributed to a characteristic included in the platform's protected categories, such as gender or gender identity. The Toolkit's guiding questions suggest that platforms establish whether a behavior is sexualized or gender-related and then assess whether gaps remain in their existing policies.

A second theme concerns the recognition of misogyny as an ideology. TikTok lists misogyny as a hateful ideology in its Hate Speech and Hateful Behavior policy, while Meta lists Nazism, white supremacy, white separatism and white nationalism in its Dangerous Organizations and Individuals policy that does not include misogyny. The Toolkit recommends that platforms recognize male-supremacist extremism as an enforceable category under existing policy frameworks, which would not require new policy architecture. Meta, for example, could add male supremacy to its existing ideology-designation list, and TikTok already names the ideology and could make the existing cross-reference from its violent extremism policy explicit.

A final theme concerns the relationship between misogyny, TFGBV and violent extremism across platform policy. Twitch's Off-Service Conduct policy includes non-consensual sexual activities, sexual assault and the sexual exploitation of youth, including grooming in its list of disqualifying conduct. These appear alongside deadly violence, violent extremism, terrorist activities and membership of a known hate group. Sexual harm and extremist affiliation are therefore treated as equivalent grounds for the same enforcement outcome. Here, the continuum between gender-based violence and extremist violence described in the Toolkit is not absent from platform policy, given that part of it is already enforced, although the policy does not name the connection or extend it to conduct occurring on the platform.

The Toolkit's framework directs platforms to escalate cases to specialist extremism or threat intelligence teams where the relevant criteria are met, following an assessment against TFGBV, hate and extremism criteria.

⑤

Recommendations

These recommendations build on ISD's Toolkit recommendations for platforms and industry but go further by drawing on the specific gaps this analysis has identified in platforms' published policies. Most can be implemented within existing policy structures, without the need for new policy architecture. The analysis is based on publicly accessible policies, which may differ from internal policy language operationalized by enforcement teams.

Narrowness is often what makes platform policies enforceable. In line with the Toolkit's emphasis on thresholds and triage rather than broadened definitions, the aim is not to make policies wider. Instead, platforms should review their existing policies, including definitions, cross-references and designation mechanisms, to determine which gender-based harms each policy is able to recognize—and how to adopt a gendered approach to policy development and internal guidance. The recommendations below set out how to do so.

01 **Protected characteristics:** extend gender-related protected characteristics to include gender expression, sex characteristics and non-adherence to prescribed gender roles or norms, and make explicit the cross-reference between hate policies' protected characteristics and violent extremism and dangerous organizations policies, so that misogyny and TFGBV are assessed across all relevant policy areas.

02 **Male-supremacist extremism as an enforceable category:** add male supremacy to existing ideology or entity designation mechanisms, such as the list of ideologies in Dangerous Organizations and Individuals policies. Explicitly link hate policies that already identify misogyny as a supremacist ideology, such as TikTok's, to violent extremism policies.

03 **Connect extremist tactics to who is targeted:** where violent extremism policies already name tactics such as harassment and doxxing, recognize that these are disproportionately directed at women and LGBTQI+ people on the basis of gender, so that coordinated campaigns by male-supremacist networks can be assessed as a pattern rather than a series of individual cases.

04 **Name the gender dimension of impersonation and account access control:** reflect in authenticity, impersonation and account-security policies that these are established TFGBV tactics, including their use for stalking, recruitment and trafficking, so that cases can be routed to the appropriate teams.

05 **Extend sexual extortion protections to adults:** where sexual extortion is addressed only within child safety policies, ensure adult victim-survivors are also explicitly covered.

06 **Apply taxonomy-based thresholds for triage and escalation:** for example, use the Toolkit's harms taxonomy and guiding questions to assess cases across TFGBV, targeted hate and violent extremism criteria, and triage to extremism or threat intelligence teams where thresholds are met.

07 **Strengthen reporting, transparency and researcher access:** map TFGBV harm categories to user reporting flows, classify harms consistently across TFGBV, hate and violent extremism in transparency reporting, and improve data access for researchers, including to assess over-removal and its consequences for freedom of expression.

Box 2 reproduces the Toolkit's guiding questions for platforms. This box is designed to support consistent, context-aware decision-making by platforms when assessing, moderating, and responding to cases involving TFGBV, targeted hate and violent extremism.

BOX 2 Guiding Questions for Platforms

STAGE	QUESTIONS AND ACTIONS
Identify the harm	Is the behavior harassment, impersonation, intimate image abuse, sexual extortion or account compromise? Is it sexualized or gender-related? Is a public figure targeted?
Check for contextual signals	Is the target a woman or LGBTQI+ person? Is the harassment coordinated or networked? Does the perpetrator reference male-supremacist or anti-LGBTQI+ narratives? Are there offline safety risks, such as doxxing, threats or stalking?
Threshold assessment	How does this case fall across TFGBV, targeted hate and violence, and violent extremism criteria?
Enforcement questions	Are there gaps in existing policies? Does this require policy review? Should evidence be preserved for law enforcement?
Systemic risk assessment	Is the case part of a broader pattern or community-wide attack? Are cross-platform networks or extremist communities involved? Is a protected group being targeted in a way that indicates structural or ongoing abuse?
Required actions	• Apply account actions (warnings, suspensions or removals); • Take appropriate content moderation action; • Hash and block future re-uploads; • Notify or support the targeted user; • Escalate to threat intelligence, extremism or specialized review teams; • Strengthen detection models where gaps are identified.

Source: ISD, Addressing the Intersection of Misogyny, Targeted Hate and Violent Extremism (Toolkit), p. 31.

Conclusion

Platforms should bridge harm categories they have traditionally kept separate, as highlighted in the Toolkit. This analysis shows both where those categories overlap and how far apart they currently sit. Taken together, the findings show that while platforms prohibit TFGBV-related behaviors, they recognize their gendered dimensions unevenly. Gender is recognized in platform policies through two main routes. The first is through protected-characteristic lists, which almost always sit within hate policies and must be referenced by other policies seeking to recognize gender-based targeting. The second is through prohibitions on conduct that is inherently sexual or consent based. However, this may capture individual incidents without recognizing the broader patterns of gender-based harm in which they occur. Although the Toolkit identifies violent extremism indicators across all five TFGBV harm types, platform policies on violent extremism and dangerous organizations explicitly address only harassment, and do so by relying on protected-characteristic definitions found in hate policies.

These findings are based on published policy language, which may differ from internal guidance used by enforcement teams. As set out in the recommendations, the implication is not necessarily that platforms need wider-ranging policies, but that they should review their existing policy structures to assess which gender-based harms each is equipped to recognize.

The annex lists all policies assessed.

Annex: Policy Inventory Table

The sample comprises 120 policy documents from 12 platforms, accessed in June 2026. These include Terms of Service, community guidelines and standards, and dedicated policy explainers. A policy document is defined as any separately named policy published by a platform, with jurisdictional versions of Terms of Service counted separately. Documents were coded at the level of the named policy or section, resulting in 121 coded entries; TikTok's US Terms of Service was coded as two sections. Because platforms divide their rules into documents differently, document counts per platform are not used as a measure of policy coverage. The full sample, including hyperlinks to the source URLs, is provided below.

PLATFORM	POLICY / LOCATION / SUB-POLICY NAME (HYPERLINKED)
TikTok	Terms of Service (EEA/UK/CH), 4.5 What you can't do on the Platform; 4.6 Your content
	Terms of Service (US) 3.4 What you can't do on the Platform
	Terms of Service (US) 3.4 What you can't do on the Platform - Section on Generative AI
	Terms of Service (If you are not in the US, EEA, the United Kingdom, Switzerland or India)
	Community Guidelines: Safety and Civility - Violent and Hateful Organizations and Individuals
	Community Guidelines: Safety and Civility - Violent and Criminal Behavior
	Community Guidelines: Safety and Civility - Hate Speech and Hateful Behavior
	Community Guidelines: Safety and Civility - Youth Sexual and Physical Abuse
	Community Guidelines: Safety and Civility - Adult Sexual Abuse
	Community Guidelines: Safety and Civility - Human Trafficking and Smuggling
	Community Guidelines: Safety and Civility - Harassment and Bullying
	Community Guidelines: Mental and Behavioral Health - Suicide and Self-Harm
	Community Guidelines: Mental and Behavioral Health - Disordered Eating and Body Image
	Community Guidelines: Mental and Behavioral Health - Dangerous Activity and Challenges
	Community Guidelines: Sensitive and Mature Themes - Body Exposure and Sexualized Behaviors
	Community Guidelines: Sensitive and Mature Themes - Shocking and Graphic Content
	Community Guidelines: Integrity and Authenticity - Misinformation
	Community Guidelines: Integrity and Authenticity - Edited Media and AI-Generated Content (AIGC)
	Community Guidelines: Integrity and Authenticity - Deceptive Behaviors and Fake Engagement
	Community Guidelines: Privacy and Security - Personal Information
	Community Guidelines: Privacy and Security - Platform Security
	Community Guidelines: Accounts and Features
X	Terms of Service (EU, EFTA States, or UK)
	Terms of Service - US, and other jurisdictions
	Rules and Policies - Violent and hateful entities policy
	Rules and Policies - Hateful Conduct
	Rules and Policies - Private Content
	Rules and Policies - Abuse and Harassment
	Rules and Policies - Adult Content
	Rules and Policies - Perpetrators of Violent Attacks
	Rules and Policies - Violent Content
	Rules and Policies - Non-consensual nudity policy
	Rules and Policies - Child Safety
	Rules and Policies - Abuse profile information

PLATFORM	POLICY / LOCATION / SUB-POLICY NAME (HYPERLINKED)
	Rules and Policies - Authenticity
YouTube	Terms of Service - GB
	Terms of Service - NL
	Terms of Service - US
	Policies - Spam & deceptive practices - Impersonation policy
	Policies - Sensitive Content - Nudity & Sexual Content Policy
	Policies - Sensitive Content - Child safety policy
	Suicide, self-harm, and eating disorders policy
	Vulgar language policy
	Violent or dangerous content - Harmful or dangerous content policy
	Violent or graphic content policies
	Violent extremist or criminal organizations policy
	Hate speech policy
	Harassment & cyberbullying policies
Twitch	Terms of Service
	Community Guidelines - Violence and Threats
	Community Guidelines - Terrorism and Violent Extremism
	Community Guidelines - Adult Sexual Violence
	Community Guidelines - Youth Safety
	Community Guidelines - Off-Service Conduct
	Community Guidelines - Unauthorized Sharing of Private Information
	Community Guidelines - Hateful Conduct
	Community Guidelines - Harassment
	Community Guidelines - Sexual Harassment
	Community Guidelines - Extreme Violence, Gore, and Other Obscene Conduct
	Community Guidelines - Sexual Content
	Community Guidelines - Account Usernames and Display Names
	Community Guidelines - Tags
	Prohibited Games
	Impersonation
	Harmful Misinformation Actors
Discord	Terms of Service
	Community Guidelines
	Bullying, Harassment, and Threats Policy Explainer
	Doxxing Policy Explainer
	Hateful Conduct Policy Explainer
	Violent Extremism Policy Explainer
	Violence and Graphic Content Policy Explainer
	Sexual Content Policy Explainer
	Off-Platform Behaviors Policy Explainer
	Non-Consensual Intimate Imagery Policy Explainer
	Identity and Authenticity Policy Explainer
	Deceptive Practices Policy Explainer
	Sexual Solicitation Policy Explainer
	Human Trafficking Policy Explainer
	Teen and Child Safety Policy Explainer
	Discord Developer Policy

PLATFORM	POLICY / LOCATION / SUB-POLICY NAME (HYPERLINKED)
Reddit	User Agreement - Reddit User Agreement if you live in the United States or outside the EEA, United Kingdom, and Switzerland
	User Agreement - Reddit User Agreement if you live in the EEA, United Kingdom, or Switzerland
	Reddit Rules
	Moderator Code of Conduct
	Harassment
	Violence
	Hate
	Manipulated Content and Misleading Behavior
	Impersonation
	How does Reddit fight the dissemination of terrorist content? *Note: this is a policy explainer / FAQ*
	Is posting someone's private or personal information okay?
	Never Post Intimate or Sexually Explicit Media of Someone Without Their Consent
	Do not share content depicting or promoting neglect, physical, or emotional abuse against minors
	Illegal or Prohibited Transactions
Rumble	Website Terms and Conditions of Use and Agency Agreement
Instagram	Instagram Terms of Use
Facebook [Meta]	Terms of Service
Facebook + Instagram [Meta]	Meta Community Standards - Coordinating harm and promoting crime
	Meta Community Standards - Dangerous organizations and individuals
	Meta Community Standards - Violence and incitement
	Meta Community Standards - Adult sexual exploitation
	Meta Community Standards - Bullying and harassment
	Meta Community Standards - Child sexual exploitation, abuse and nudity
	Meta Community Standards - Human Exploitation
	Meta Community Standards - Suicide, self-injury and eating disorders
	Meta Community Standards - Hateful Conduct
	Meta Community Standards - Privacy Violations
	Meta Community Standards - Violent and Graphic Content
	Meta Community Standards - Account Integrity
	Meta Community Standards - Authentic identity representation
	Meta Community Standards - Inauthentic Behavior
	Meta Community Standards - Cybersecurity
	Meta Community Standards - Adult Sexual Solicitation and Sexually Explicit Language
	Meta Community Standards - Fraud, scams and deceptive practices
	Meta Community Standards - Adult nudity and sexual activity
	Meta Community Standards - Misinformation
Telegram	Terms of Service
	User guidance for the EU Digital Services Act
4chan	Rules - Global Rules
Incels.is	Terms and Rules



Amman | Berlin | London | Paris | Toronto | Washington DC

Copyright © Institute for Strategic Dialogue (2026). Institute for Strategic Dialogue (ISD) is a company limited by guarantee, registered office address 3rd Floor, 45 Albemarle Street, Mayfair, London, W1S 4JL. ISD is registered in England with company registration number 06581421 and registered charity number 1141069. All Rights Reserved.

ISD | Institute for Strategic Dialogue

Christchurch Call Foundation