doi: 10.21783/rei.v12i2.1031

SUBMETIDO: 20 MAIO 2026
ACEITO: 20 MAIO 2026

ALGORITMOS COMO INSTITUIÇÕES DA VIOLÊNCIA DE GÊNERO ONLINE: PLATAFORMAS DIGITAIS, OBRIGAÇÕES EMPRESARIAIS DE DIREITOS HUMANOS E RESPONSABILIDADE INTERNACIONAL DO ESTADO

ALGORITHMS AS INSTITUTIONS OF ONLINE GENDER-BASED VIOLENCE: DIGITAL PLATFORMS, BUSINESS AND HUMAN RIGHTS OBLIGATIONS AND INTERNATIONAL STATE RESPONSIBILITY

YASMIN CURZI DE MENDONÇA [1]

RESUMO: Este artigo procura compreender quais as obrigações de devida diligência para empresas e para o Estado brasileiro à luz do direito internacional dos direitos humanos em relação à violência de gênero online nas plataformas de redes sociais. A metodologia empregada é revisão de literatura de governança de plataformas e da teoria institucional dos algoritmos, além de análise do arcabouço normativo internacional, com ênfase nas obrigações derivadas da Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW), da Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher (Convenção de Belém do Pará) e dos Princípios Orientadores sobre Empresas e Direitos Humanos da ONU (Princípios Ruggie). O artigo argumenta

[1] Pesquisadora de pós-doutorado no Karsh Institute of Democracy na University of Virginia. Professora na FGV Direito Rio, onde coordena o Programa de Diversidade e Inclusão e atua como pesquisadora no Centro de Tecnologia e Sociedade (CTS-FGV) desde 2019, com foco em Direitos Humanos, Tecnologia, Regulação de Plataformas, Gênero e Democracia Digital e Cibersegurança. Doutora pelo IESP-UERJ, cursado com bolsa CAPES e indicação à bolsa FAPERJ Nota 10 por desempenho acadêmico. Mestre pelo PPGCIS da PUC-Rio, com bolsa CAPES. Graduada em Direito e em Ciências Sociais pela Fundação Getulio Vargas, com período de intercâmbio acadêmico na Université Sorbonne Paris-IV. Atualmente, coordena na FGV Direito Rio o projeto "Digital Media and Conflict Prevention" (2023-2025), financiado por uma Horizon Grant da União Europeia, focado em desinformação de gênero e ataques contra mulheres na política, jornalistas e ativistas online. Anteriormente, atuou na ONG Stop Street Harassment no Brasil e no Center for Justice and International Law (CEJIL). Advogada pela OAB-RJ e vice-presidente da Comissão de Liberdade de Expressão no Mundo Digital.

que sistemas algorítmicos não são ferramentas neutras, mas dispositivos que potencializam a disseminação e monetização da violência de gênero de forma estrutural, e conclui que o Estado brasileiro pode ser responsabilizado pela omissão regulatória que permite a persistência de tais violações.

PALAVRAS-CHAVE: violência de gênero online; algoritmos como instituições; empresas e direitos humanos; tecno-silenciamentos; CEDAW; Convenção de Belém do Pará; teoria da atribuição.

ABSTRACT: This article seeks to understand what due diligence obligations apply to companies and to the Brazilian state under international human rights law in relation to online gender-based violence on social media platforms. The methodology employed consists of a literature review of platform governance scholarship and the institutional theory of algorithms, combined with an analysis of the international normative framework, with emphasis on obligations derived from Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), the Inter-American Convention on the Prevention, Punishment, and Eradication of Violence against Women (Belém do Pará Convention), and the UN Guiding Principles on Business and Human Rights (Ruggie Principles). The article argues that algorithmic systems are not neutral tools, but devices that structurally enable the dissemination and monetization of gender-based violence and concludes that the Brazilian state may be held accountable for the regulatory omission that allows such violations to persist.

KEYWORDS: online gender-based violence; algorithms as institutions; business and human rights; techno-silencings; CEDAW; Belém do Pará Convention; theory of attribution.

INTRODUÇÃO

A Violência de Gênero Online (VGO) tem crescido substancialmente no Brasil. Dados do SaferNet em parceria com o Observatório Nacional dos Direitos Humanos (BRASIL, 2024) indicam que, entre 2017 e 2022, denúncias de discursos de ódio contra mulheres na internet aumentaram de 961 para 28,6 mil registros, crescimento de quase trinta vezes em cinco anos. Uma pesquisa empírica do NetLab UFRJ, conduzida em parceria com o Ministério das Mulheres em 2024, identificou 137 canais do YouTube brasileiro com conteúdo misógino que, somados, acumulavam 3,9 bilhões de visualizações; 80% desses canais possuíam ao menos uma estratégia ativa de monetização (Santini et al., 2024). Em relação a mulheres na política, monitoramento do projeto MonitorA revelou que 11% dos tuítes dirigidos a candidatas nas eleições municipais de 2020 continham linguagem

misógina, racista ou ofensiva, com ataques focados não em posições políticas, mas em gênero, raça e moralidade (AzMina; InternetLab, 2020).

A escala e a sistematicidade desses dados conduzem a uma questão central: quais seriam as obrigações de devida diligência para empresas e para o Estado brasileiro à luz do direito internacional dos direitos humanos em relação à VGO? Para respondê-la, o artigo parte de uma hipótese analítica derivada do *framework* do institucionalismo algorítmico (Mendonça; Filgueiras; Almeida, 2025): os algoritmos das plataformas digitais funcionam como instituições da VGO no Brasil, na medida em que organizam identidades e estabelecem comportamentos de maneira que tende a reproduzir e amplificar hierarquias de gênero preexistentes. A metodologia combina revisão de literatura de governança de plataformas e da teoria institucional dos algoritmos com análise do arcabouço normativo internacional.

O artigo analisa, no primeiro plano, as obrigações empresariais derivadas dos Princípios Orientadores sobre Empresas e Direitos Humanos da ONU (Princípios Ruggie, 2011), que estabelecem responsabilidade autônoma das empresas para conduzir medidas de devida diligência sobre impactos de suas operações, incluindo, como a literatura recente examina, o design algorítmico e a moderação de conteúdo (Suzor et al., 2019; Klonick, 2018; Gillespie, 2018; Grimmelmann, 2015). No segundo plano, examina a teoria da atribuição no direito internacional (CDI, 2001), articulada com as obrigações de devida diligência derivadas da Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW) e da Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher (Convenção de Belém do Pará), como fundamento para avaliar a omissão regulatória do Estado brasileiro.

O argumento se desenvolve em três partes. A primeira examina os fundamentos teóricos do institucionalismo algorítmico e suas implicações para a análise jurídica da VGO. Situa-se também a VGO como fenômeno sistêmico no Brasil, à luz dos dados disponíveis sobre a monetização de seu espraiamento. A segunda analisa as obrigações internacionais pertinentes e sua articulação com a teoria da atribuição. A terceira examina os mecanismos de tecno-silenciamento (Curzi, 2026) nas plataformas e os limites da autorregulação empresarial. A conclusão sistematiza os achados e aponta questões em aberto.

2. ALGORITMOS COMO INSTITUIÇÕES: FUNDAMENTOS TEÓRICOS

A literatura de governança de plataformas tem progressivamente questionado a representação das plataformas digitais como infraestruturas técnicas neutras. Desde os trabalhos pioneiros de Joel Reidenberg (1997) sobre a *lex informatica* e de Lawrence Lessig (2000) sobre a arquitetura do ciberespaço como forma de regulação, passando pelos estudos sobre plataformização da *web* (Helmond, 2019) e moderação de conteúdo como exercício de poder normativo (Goldman, 2021; Gillespie, 2018; Klonick, 2018; Grimmelmann, 2015), acumulou-se evidência de que

plataformas digitais elaboram políticas de uso, moderam conflitos, ponderam direitos e sancionam usuários por meio de sistemas algorítmicos e humanos cuja transparência, previsibilidade e condições de devido processo são limitadas (Hartmann, et al. 2026).

Nesse cenário, a questão analítica que se impõe é em que medida esse poder normativo pode ser compreendido como institucional. Como desenvolvido por Mendonça, Filgueiras e Almeida (2025), algoritmos funcionam como instituições políticas na medida em que configuram relações de poder, organizam identidades e estabelecem comportamentos apropriados em diversas arenas de interação social, segundo uma lógica de adequação análoga à descrita por March e Olsen (2009) para as instituições políticas tradicionais. O poder dos algoritmos, assim, manifesta-se na capacidade de definir os indivíduos e situá-los em posições pré-configuradas. Sujeitos tornam-se pontos de dados classificados em clusters que retroagem sobre seus comportamentos, tendendo a cristalizar e reforçar hierarquias existentes.

Para os fins deste artigo, tal lógica de fixação identitária é relevante por uma razão específica: em ambientes sociais em que a hostilidade de gênero é culturalmente disponível e, como se examinará na seção seguinte, economicamente rentável para as plataformas, esses sistemas podem produzir resultados sistematicamente desfavoráveis para mulheres e pessoas de gênero dissidente.

Como argumentam Krupiy e Scheinin (2023), sistemas algorítmicos produzem danos não por referências explícitas a características protegidas como gênero ou raça, mas por meio de outras variáveis, como as lógicas de otimização e outros efeitos sistêmicos emergentes. Essa discriminação digital difere da discriminação direta e indireta tradicionais, pois opera por correlações opacas entre pontos de dados e representações abstratas de indivíduos dentro de modelos algorítmicos, frequentemente reforçando desigualdades estruturais enquanto aparece como neutra. Tais características desafiam os marcos jurídicos tradicionais e demandam desenvolvimento doutrinário no plano do direito internacional dos direitos humanos. A análise da VGO nas plataformas brasileiras exige, portanto, considerar o contexto social em que esses sistemas operam.

2.1. VGO NO BRASIL: CAPITALISMO DE PLATAFORMAS E TECNO-SILENCIAMENTOS

O modelo de negócios das plataformas digitais, baseado na maximização do engajamento por meio da monetização da atenção, é ponto de partida analítico relevante para compreender a VGO. A literatura sobre economia política das plataformas sugere que algoritmos de recomendação tendem a privilegiar conteúdos que evocam indignação, ressentimento e hostilidade, por gerarem maior engajamento (Munger, 2024; Marwick, 2021; Zuboff, 2019). Soma-se a isso o fato de que, como aponta Franks (2019), pouca atenção foi dada ao efeito silenciador de ataques online que afetam desproporcionalmente grupos vulneráveis. Por décadas, o discurso dominante sobre "*chilling effects*" nas grandes plataformas tem focado sistematicamente na ameaça de supressão indevida de conteúdo, enquanto ignora

o silenciamento estrutural de vozes marginalizadas pelo próprio abuso que as plataformas permitem circular (Curzi, 2026).

Como demonstrado por uma pesquisa empírica conduzida pelo NetLab UFRJ, essa dinâmica se apresenta de forma substantiva no YouTube brasileiro (Santini et al., 2024). A partir de um *corpus* de 76,3 mil vídeos de 7,8 mil canais, o estudo identificou 137 canais com conteúdo inequivocamente misógino. O tema predominante no *corpus*, presente em 42% dos vídeos, foi categorizado como "desprezo às mulheres e estímulo à insurgência masculina", i.e., conteúdos que promovem a crença em uma suposta conspiração pela dominação feminina e incitam resistência a ela. O volume de conteúdos cresceu expressivamente a partir de 2022; 88% do *corpus* havia sido publicado desde 2021. A análise da série histórica sugere que esse crescimento acompanha o avanço de perspectivas neoconservadoras articuladas à polarização política do período, fenômeno que Faludi (2001) conceituou como *backlash*, ou seja, um contra-ataque para impedir o progresso das mulheres a partir de uma lógica de ressentimento.

Em relação às práticas de monetização, dos 137 canais misóginos identificados pelo NetLab, 80% possuíam ao menos uma estratégia ativa de geração de receita: 52% geravam renda por meio de anúncios veiculados pela própria plataforma, 28% pelo programa de membros do YouTube, e oito canais arrecadaram mais de R$ 68 mil em doações em transmissões ao vivo. Além dos mecanismos nativos, os canais utilizavam plataformas externas como Hotmart e Kiwify para vender cursos, e-books e consultorias individuais de até R$ 1 mil (Santini et al., 2024). De tal maneira, a inação corporativa diante da VGO é potencialmente integrativa de um modelo de negócios em que o engajamento gerado por conteúdos de ódio é economicamente rentável tanto para os produtores quanto para a própria plataforma.

A "machosfera" brasileira, ecossistema digital de comunidades masculinistas composto por canais, páginas e fóruns que abordam masculinidade, relacionamentos e crítica ao feminismo (Ging, 2017; Curzi, 2026), apresenta quatro subculturas predominantes, como também identificado pela pesquisa do NetLab: Red Pill, MGTOW (Men Going Their Own Way), Pick Up Artists e celibatários involuntários (incel). Seus discursos compartilham a premissa de que os homens são vítimas de uma ordem social feminocêntrica e que devem resistir a ela por meio do controle, desprezo ou afastamento das mulheres (Santini et al., 2024; Han; Yin, 2023). Seguindo a análise da filósofa Kate Manne (2018), tais narrativas operam como sistema de controle social que policia e impõe a ordem patriarcal, punindo as mulheres que desafiam papéis de gênero estabelecidos.

Nesse sentido, as plataformas digitais têm oferecido uma infraestrutura pela qual esse controle social se articula, se amplifica e se monetiza (Curzi, 2026). A exclusão estrutural de mulheres e outros grupos marginalizados das esferas públicas digitais é viabilizada pela interação entre falhas de governança de plataformas e omissão regulatória estatal persistente. Tais dinâmicas constituem, como observaram as relatoras especiais da ONU (Khan, 2021; Simonovic, 2018) e a literatura sobre

padrões de resistência ao protagonismo feminino no espaço digital (Sobieraj, 2017), uma barreira estrutural à igualdade com consequências para o exercício do direito à liberdade de expressão e participação no debate público.

3. OBRIGAÇÕES DERIVADAS DO DIREITO INTERNACIONAL DE DIREITOS HUMANOS

3.1. O ALCANCE DOS PRINCÍPIOS RUGGIE

Os Princípios Orientadores sobre Empresas e Direitos Humanos da ONU (Princípios Ruggie, 2011) estabelecem o marco "Proteger, Respeitar e Reparar": os Estados têm o dever de proteger contra violações cometidas por terceiros; as empresas têm responsabilidade autônoma de respeitar direitos humanos; e ambos devem garantir acesso a mecanismos de reparação. Como observam Suzor et al. (2019), os *frameworks* de direitos humanos foram concebidos originalmente com foco em atores estatais, enraizados na divisão entre público e privado que excluía historicamente as corporações de obrigações diretas sob o direito internacional. Os Princípios Ruggie representam uma inflexão nesse paradigma ao articular que empresas privadas têm responsabilidade de evitar infringir direitos humanos e remediar impactos adversos com os quais estejam envolvidas.

A responsabilidade empresarial de respeitar direitos humanos inclui o dever de conduzir medidas de devida diligência para identificar, prevenir e mitigar impactos adversos decorrentes de suas operações (Princípio 17). O conteúdo concreto dessa devida diligência para empresas de plataformas, entretanto, ainda é objeto de debate. Uma interpretação extensiva sustentaria que o *design* algorítmico, por ser a operação central dessas empresas e por produzir impactos documentados sobre populações vulnerabilizadas, deveria ser objeto de avaliação periódica de impacto em direitos humanos com recorte de gênero e raça. Uma interpretação mais restritiva, contudo, apontaria que a natureza de *soft law* dos Princípios Ruggie não traz obrigações detalhadas e específicas. O cenário fático, portanto, permite que empresas se apresentem como em *compliance* em direitos humanos enquanto retêm plena discricionariedade sobre quando, como e se devem agir (Suzor et al., 2019). A validação desse modelo teria, assim, produzido a institucionalização de um cumprimento meramente performático. Como apontado por Caplan (2023), empresas têm implementado relatórios de transparência e medidas de governança aparentemente descentralizadas, como com conselhos de segurança, ou parcerias com organizações da sociedade civil e especialistas, mas sem endereçar adequadamente as infraestruturas projetadas para viralidade e lucro.

No entanto, esse déficit de *accountability* não decorre apenas da arquitetura voluntarista dos Princípios Ruggie, mas também da insuficiência regulatória dos Estados. O primeiro pilar do marco, i.e., o dever estatal de proteger (Princípio 1), impõe aos Estados a obrigação de adotar medidas legislativas, administrativas e judiciais para prevenir, investigar, punir e reparar violações cometidas por empresas em sua jurisdição, o que, no contexto da VGO, implica obrigação

afirmativa de criar condições normativas que tornem o *compliance* substantivo, e não apenas declaratório.

O alcance da responsabilidade empresarial, nos termos dos Princípios Ruggie, é, portanto, relevante para a análise de seu papel na amplificação da VGO como problema de atribuição compartilhada entre atores privados e Estados. Nas plataformas digitais, essa responsabilidade se estende além das interações diretas com usuários, abrangendo o ecossistema de desenvolvedores, anunciantes, moderadores e autoridades públicas. A questão jurídica relevante é se a contribuição algorítmica para a amplificação de conteúdos de ódio de gênero deveria ser enquadrada como impacto adverso nos termos dos Princípios Ruggie. A evidência examinada na seção seguinte sugere que essa contribuição existe e é mensurável, embora o nexo causal entre design algorítmico específico e danos concretos permaneça como questão empírica a ser desenvolvida.

3.2. DEVIDA DILIGÊNCIA NA CEDAW E CONVENÇÃO DE BELÉM DO PARÁ

A Recomendação Geral nº 19 do Comitê CEDAW (1992) afirmou que a violência baseada no gênero constitui forma de discriminação nos termos da Convenção. Enquanto a Convenção em si não menciona explicitamente a violência, focando apenas na discriminação, o Comitê expandiu progressivamente o escopo das obrigações estatais por meio de seus instrumentos interpretativos. A Recomendação Geral nº 35 (2017) atualizou esse entendimento para incluir explicitamente a violência mediada por tecnologias de informação e comunicação, reconhecendo que a violência de gênero pode ser frequentemente exacerbada por fatores tecnológicos, culturais, ideológicos e econômicos, e exigindo que os Estados monitorem e regulem as atividades de atores privados cuja conduta possa perpetuar a discriminação de gênero. Mais especificamente, o Comitê afirma que, como parte de suas obrigações de devida diligência, os Estados devem adotar e implementar diversas medidas para lidar com a violência de gênero cometida por atores não estatais, o que inclui a adoção de leis e de um sistema em vigor para lidar com tal violência.

Krupiy (2021) argumenta que, embora a CEDAW ofereça base normativa robusta para abordar discriminações de gênero estruturais, ela requer desenvolvimento doutrinário e interpretativo para responder à dimensão digital de forma substantiva. Sistemas algorítmicos e de inteligência artificial produzem dano não por referências explícitas a características protegidas como gênero ou raça, mas por operação de outras variáveis, como as lógicas de otimização e seus efeitos sistêmicos emergentes. Assim, a discriminação digital opera por correlações opacas entre pontos de dados, frequentemente reforçando desigualdades estruturais enquanto seus sistemas tendem a parecer neutros aos espectadores e usuários, o que desafia os marcos jurídicos tradicionais de discriminação direta e indireta e demanda, como propõem Krupiy e Scheinin (2023), o reconhecimento de uma "obrigação positiva" dos Estados de identificar e prevenir impactos

discriminatórios, especialmente quando design digital e omissão estatal se intersectam com desigualdades estruturais.

Adicionalmente, o relatório da Relatora Especial sobre Violência contra a Mulher, Dubravka Simonovic (A/HRC/38/47, 2018), consolidou, no âmbito da ONU, uma definição para violência online contra mulheres que abrange qualquer ato cometido, assistido ou agravado por tecnologias que produza danos psicológicos, físicos, sexuais e econômicos, silenciamento e retração da vida pública. O relatório reconheceu que as características do ambiente digital, como a viralidade, persistência, escalabilidade e anonimato, amplificam as violências. Além disso, em 2021, a Relatora Especial sobre Liberdade de Expressão, Irene Khan, fez contribuição significativa ao identificar e examinar explicitamente a "desinformação generificada" como tipo de violência de gênero, expondo as formas estruturais, sistêmicas e simbólicas de censura generificada que impedem a participação igualitária de mulheres e pessoas de gênero diverso na esfera pública (Khan, 2021). O relatório articulou como essa censura opera não apenas pela repressão estatal direta, mas também, e criticamente, pelas ações e omissões de atores privados, incluindo plataformas digitais, que silenciam vozes femininas, negam validade às suas experiências e as excluem do discurso público.

No âmbito do Sistema Interamericano de Direitos Humanos, a Convenção de Belém do Pará (1994), embora elaborada em contexto pré-internet, tem sido interpretada de forma evolutiva pelo Mecanismo de Seguimento da Convenção de Belém do Pará (MESECVI) e pela Comissão Interamericana de Direitos Humanos (CIDH) como aplicável à violência baseada em gênero no ambiente digital. O MESECVI tem sublinhado que os Estados devem adotar marcos legais e institucionais capazes de abordar a natureza evolutiva da violência contra as mulheres em ambientes digitais, o que culminou no processo de elaboração da Lei Modelo Interamericana para Prevenir, Punir e Erradicar a Violência Digital de Gênero contra as Mulheres (Curzi, 2026). De tal modo, a omissão estatal em regular plataformas, prover remédios ou agir contra a cumplicidade do setor privado em tais danos poderia constituir violação de obrigações vinculantes.

A jurisprudência da Corte Interamericana contribui com princípios cujo alcance ao contexto digital merece exame. Em *González e outras vs. México* (2009), a Corte reconheceu a violência contra mulheres como problema estrutural e estabeleceu obrigação reforçada de prevenção com base no artigo 7 da Convenção de Belém do Pará. Em *Bedoya Lima vs. Colômbia* (2021), a Corte destacou que mulheres jornalistas enfrentam riscos acrescidos pela combinação de gênero e atuação profissional, padrão documentado também no contexto brasileiro (ABRAJI, 2021), enfatizando o dever estatal de proteção especial. A questão analítica é se, e em que medida, esses precedentes, desenvolvidos em contextos de violência física com nexo causal mais direto com a omissão estatal, podem ser estendidos ao contexto da violência algorítmica mediada por plataformas privadas.

3.3. TEORIA DA ATRIBUIÇÃO E RESPONSABILIDADE ESTATAL POR OMISSÃO REGULATÓRIA

O princípio da devida diligência impõe ao Estado a obrigação de adotarem medidas razoáveis para prevenir, investigar e reparar atos de violência de gênero praticados por particulares. A articulação técnica dessa obrigação com a conduta de atores privados como as plataformas exige o recurso à teoria da atribuição no direito internacional, consolidada pelos Artigos sobre Responsabilidade do Estado por Atos Internacionalmente Ilícitos (CDI, 2001). O artigo 2 estabelece que um ato internacionalmente ilícito existe quando há conduta atribuível ao Estado que constitua violação de uma obrigação internacional.

A jurisprudência interamericana avançou para reconhecer a atribuição por omissão como fundamento autônomo de responsabilidade. No caso Maria da Penha, a CIDH (Relatório de Mérito nº 54/01, 2001) concluiu que o Brasil violou as obrigações da Convenção de Belém do Pará pela tolerância da situação de violência doméstica, configurando um padrão de negligência e ineficácia institucional. Contudo, os casos em que essa doutrina foi desenvolvida envolviam situações em que o Estado tinha conhecimento de risco concreto para vítimas identificáveis. A VGO apresenta características distintas: os danos são estruturais, difusos e mediados por sistemas técnicos operados por empresas privadas globais frequentemente situadas fora da jurisdição do Estado. A questão é quais seriam os critérios para determinar que o Estado dispunha de conhecimento suficiente do risco para que sua inação configure violação de obrigação convencional.

Em um contexto em que a escala e a recorrência da VGO online estão bem documentadas, a inércia continuada do Estado brasileiro não é uma questão de capacidade apenas, mas consequência da ausência estrutural de políticas para abordar a VGO em suas raízes (Curzi, 2026). No plano da omissão direta, quando o Estado, diante de evidências de que os algoritmos das plataformas produzem sistematicamente VGO, deixa de adotar as medidas regulatórias exigidas pela CEDAW e pela Convenção de Belém do Pará, essa omissão pode ser atribuível ao Estado como violação de suas obrigações convencionais.

O artigo 19 do Marco Civil da Internet (Lei nº 12.965/2014) é o elemento central para examinar o plano da cumplicidade estrutural. Ao condicionar a responsabilidade civil das plataformas por danos de conteúdo de terceiros à existência de ordem judicial específica de remoção, o dispositivo criou, segundo seus críticos, desincentivos à ação proativa das plataformas e impôs às vítimas o ônus de custear procedimentos judiciais lentos e onerosos (Dias et al., 2023). Em junho de 2025, o Supremo Tribunal Federal, nos julgamentos dos Temas 533 e 987, declarou o artigo 19 parcialmente e progressivamente inconstitucional, estabelecendo que as plataformas podem ser responsabilizadas civilmente mesmo sem ordem judicial prévia em casos de conteúdo pago ou artificialmente amplificado, ou quando há disseminação massiva de material ilegal grave, incluindo discurso de ódio e violência de gênero online. A decisão reconhece explicitamente um estado de omissão legislativa parcial e exorta o Congresso a

elaborar novo marco legal, representando alinhamento parcial com as obrigações internacionais de direitos humanos para prevenir a VGO (Curzi, 2026).

4. MECANISMOS DE TECNO-SILENCIAMENTO E LIMITES DA AUTORREGULAÇÃO ALGORÍTMICA

A moderação de conteúdo, como observa Gillespie (2018), é uma prática que define as plataformas enquanto categoria, mas ela é exercida com discricionariedade opaca, resultados inconsistentes e resistência à *accountability*, particularmente no Sul Global e em jurisdições não anglófonas. Estudo da Global Witness e da Internet Freedom Foundation (2024) ilustra essa dinâmica: os pesquisadores reportaram ao YouTube 79 vídeos com discurso de ódio contra mulheres para acompanhar a resposta da empresa. O YouTube confirmou o recebimento, mas apenas um dos vídeos teve restrições mínimas aplicadas. Esse dado é coerente com a crítica de Citron e Franks (2020) de que a falha das plataformas em implementar estratégias efetivas de moderação e sua relutância em remover conteúdo prejudicial são intencionais. De tal maneira, elas seguem aprofundando de forma inconsequente desigualdades estruturais.

A regressão regulatória observada entre as principais plataformas a partir de 2025 acrescenta uma dimensão adicional ao problema. Após a segunda eleição de Donald Trump nos Estados Unidos, Meta, X (antigo Twitter) e YouTube reduziram suas equipes de moderação, enfraqueceram políticas de conteúdo e retrocederam em medidas de proteção anteriormente adotadas, invocando argumentos sobre liberdade de expressão que espelham o modelo norte-americano de proteção quase absoluta ao discurso. Esse retrocesso é problemático do ponto de vista do direito internacional dos direitos humanos, que, como observa a Relatora Khan (2021), requer uma abordagem ponderada que considere tanto a liberdade de expressão quanto a proteção de grupos vulneráveis, e que trate o discurso misógino, quando leva a dano real e iminente, como discurso potencialmente restringível nos termos dos artigos 19(3) e 20(2) do PIDCP.

Para sanar esse cenário, a Lei Modelo Interamericana em elaboração pelo MESECVI propõe obrigações específicas para plataformas, como a remoção de conteúdos de VGO em até 48 horas após notificação das vítimas, sob pena de responsabilidade civil, bem como obrigações de transparência sobre moderação, relatórios periódicos e auditoria algorítmica com recorte de gênero. Essas propostas respondem diretamente a algumas das lacunas identificadas neste artigo e constituem um desenvolvimento normativo relevante para o debate sobre a efetividade dos mecanismos de reparação disponíveis. Contudo, como observa a literatura sobre a Lei de Serviços Digitais europeia (Enarsson, 2024), mesmo obrigações legalmente vinculantes enfrentam desafios de implementação quando os sistemas a regular são opacos por design e quando a capacidade regulatória dos Estados é assimétrica em relação às grandes empresas de tecnologia.

5. CONCLUSÃO

A análise desenvolvida nas seções anteriores permite sistematizar três eixos de conclusão, mantendo a distinção entre o que a evidência disponível permite afirmar e as questões que permanecem em aberto. Primeiro, no plano teórico: a hipótese do institucionalismo algorítmico (Mendonça; Filgueiras; Almeida, 2025) oferece um quadro analítico plausível para compreender os algoritmos das plataformas digitais como dispositivos que tendem a reproduzir e amplificar hierarquias de gênero preexistentes. Os dados sobre aumento de ocorrências e monetização da misoginia, além das estratégias de evasão da moderação de conteúdo por empresas são consistentes com essa hipótese, mas não a demonstram de forma conclusiva. O nexo causal específico entre design algorítmico e danos de gênero documentados permanece como questão empírica que requer pesquisa adicional, especialmente sobre os sistemas de recomendação do YouTube e os mecanismos pelos quais conteúdos da manosfera ganharam a escala observada.

Segundo, no plano das obrigações empresariais, os Princípios Ruggie (2011) fornecem base normativa para argumentar que as plataformas têm responsabilidade autônoma de conduzir medidas de devida diligência sobre os impactos de seu design algorítmico em relação às populações vulnerabilizadas. A evidência disponível, particularmente sobre a monetização de conteúdo misógino pela própria infraestrutura da plataforma e sobre os padrões sistemáticos de moderação ineficaz documentados por diversos autores como Gillespie (2018), Citron e Franks (2020), Curzi (2026), entre outros, sugere que essa responsabilidade não está sendo cumprida. A crítica de Suzor et al. (2019) sobre o cumprimento performático dos Princípios Ruggie aponta para a necessidade de mecanismos mais robustos de *accountability*, como os propostos pela Lei Modelo Interamericana do MESECVI.

Terceiro, no plano da responsabilidade estatal, a teoria da atribuição por omissão, articulada com o princípio da devida diligência derivado da CEDAW (Recomendações Gerais 19 e 35) e da Convenção de Belém do Pará, oferece fundamento normativo para avaliar a responsabilidade internacional do Estado brasileiro por sua omissão regulatória em relação às plataformas. O julgamento dos Temas 987 e 533 pelo STF (2025) representa alinhamento parcial com essas obrigações, mas a ausência de legislação abrangente e mecanismos efetivos de enforcement mantém a proteção de direitos fundamentais no ambiente digital precária e desigualmente aplicada. O padrão de omissão institucional documentado nos dados examinados sugere que a falha do Estado não decorre apenas de limitações de capacidade, mas de ausência estrutural de políticas para enfrentar a VGO em suas dimensões algorítmicas.

REFERÊNCIAS

ABRAJI – Associação Brasileira de Jornalismo Investigativo. Violência de gênero contra jornalistas. São Paulo: ABRAJI, 2021.

AZMINA; INTERNETLAB. MonitorA: relatório sobre violência política online em páginas e perfis de candidatas(os) nas eleições municipais de 2020. São Paulo, 2020.

BRASIL. Ministério dos Direitos Humanos e da Cidadania (MDHC). Incitação à violência contra a vida na internet lidera violações de direitos humanos com mais de 76 mil casos em cinco anos, aponta ObservaDH. Brasília: MDHC, 23 jan. 2024. Disponível em: https://www.gov.br/mdh/pt-br/assuntos/noticias/2024/janeiro/incitacao-a-violencia-contra-a-vida-na-internet-lidera-violacoes-de-direitos-humanos-com-mais-de-76-mil-casos-em-cinco-anos-aponta-observadh. Acesso em: 9 mar. 2026.

CAPLAN, Robyn. Networked platform governance: the construction of the democratic platform. International Journal of Communication, Los Angeles, v. 17, p. 1–22, 2023.

CIDH – Comissão Interamericana de Direitos Humanos. Relatório de Mérito nº 54/01, Caso 12.051 (Maria da Penha Maia Fernandes vs. Brasil). 2001.

CITRON, Danielle Keats. Hate crimes in cyberspace. Cambridge: Harvard University Press, 2014.

CITRON, Danielle Keats; FRANKS, Mary Anne. The Internet as a speech machine and other myths confounding Section 230 speech reform. University of Chicago Legal Forum, Chicago, v. 2020, p. 45–75, 2020.

COMISSÃO DE DIREITO INTERNACIONAL (CDI). Artigos sobre Responsabilidade do Estado por Atos Internacionalmente Ilícitos. Nova York: ONU, 2001. Disponível em: https://legal.un.org/ilc/texts/instruments/english/draft_articles/9_6_2001.pdf. Acesso em: 9 mar. 2026.

COMITÊ CEDAW. Recomendação Geral nº 19 sobre violência contra as mulheres. Nova York: ONU, 1992.

COMITÊ CEDAW. Recomendação Geral nº 35 sobre violência contra as mulheres baseada no gênero, atualizando a Recomendação Geral nº 19. Nova York: ONU, 2017.

CORTE INTERAMERICANA DE DIREITOS HUMANOS. González e outras ("Campo Algodonero") vs. México. Sentença de 16 de novembro de 2009.

CORTE INTERAMERICANA DE DIREITOS HUMANOS. Bedoya Lima vs. Colômbia. Sentença de 26 de agosto de 2021.

CURZI, Yasmin. Violência de Gênero Online: Regulação de Plataformas e Proteção de Direitos Humanos de Mulheres e Pessoas de Gênero Dissidente no Ambiente Digital. Rio de Janeiro: Lumen Juris, 2026.

DIAS, Daniel Pereira Nunes; BELLI, Luca; ZINGALES, Nicolo; GASPAR, Walkiria; CURZI, Yasmin. Plataformas no Marco Civil da Internet: a necessidade de uma responsabilidade progressiva baseada em riscos. Civilistica.com, Rio de Janeiro, v. 12, n. 3, p. 1–24, 2023.

ENARSSON, Therese. Navigating hate speech and content moderation under the DSA: insights from ECtHR case law. Information & Communications Technology Law, v. 33, n. 3, p. 384–401, 2024. Disponível em: https://www.tandfonline.com/doi/full/10.1080/13600834.2024.2395579. Acesso em: 9 mar. 2026.

FALUDI, Susan. Backlash: o contra-ataque na guerra não declarada contra as mulheres. Rio de Janeiro: Rocco, 2001.

FORTUNA, Paula; NUNES, Sérgio. A survey on automatic detection of hate speech in text. ACM Computing Surveys, v. 51, n. 4, p. 1–30, 2019. Disponível em: https://doi.org/10.1145/3232676. Acesso em: 9 mar. 2026

FRANKS, Mary Anne. The cult of the constitution. Stanford: Stanford University Press, 2019.

GILLESPIE, Tarleton. Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media. New Haven: Yale University Press, 2018.

GING, Debbie. Alphas, Betas, and Incels: theorizing the masculinities of the manosphere. Men and Masculinities, v. 22, n. 4, p. 638–657, 2017. Disponível em: https://doi.org/10.1177/1097184X177064. Acesso em: 9 mar. 2026.

GLOBAL WITNESS; INTERNET FREEDOM FOUNDATION. Letting hate flourish: YouTube and Koo's lax response to the reporting of hate speech against women. 2024. Disponível em: https://globalwitness.org/en/campaigns/digital-

threats/letting-hate-flourish-youtube-and-koos-lax-response-to-the-reporting-of-hate-speech-against-women-in-india-and-the-us/. Acesso em: 9 mar. 2026.

GOLDMAN, Eric. Content moderation remedies. Michigan Technology Law Review, v. 28, n. 1, 2021. Disponível em: http://dx.doi.org/10.2139/ssrn.3810580. Acesso em: 9 mar. 2026.

GRIMMELMANN, James. The virtues of moderation. Yale Journal of Law and Technology, v. 17, n. 1, 2015. Disponível em: https://scholarship.law.cornell.edu/facpub/1486/. Acesso em: 9 mar. 2026.

HAN, Xiaoting; YIN, Chenjun. Mapping the manosphere: categorization of reactionary masculinity discourses in digital environment. Feminist Media Studies, v. 23, n. 5, p. 1923–1940, 2023. Disponível em: https://xyonline.net/sites/xyonline.net/files/2024-06/Han%2C%20Mapping%20the%20manosphere%202023.pdf. Acesso em: 9 mar. 2026.

HARTMANN, I.; COSTA, R.; CRUZ, F. B.; KIRA, B. Dever de Cuidado de Plataformas após a Decisão do Supremo Tribunal Federal sobre o Marco Civil da InternetUniversity of Sussex, , 16 jan. 2026. Disponível em: . Acesso em: 17 maio. 2026

HELMOND, Anne. A plataformização da web. In: Omena, J. J. Métodos Digitais: Teoria-Prática-Crítica. Lisboa: Instituto de Comunicação da Nova, 2019. Disponível em: https://dspace.library.uu.nl/bitstream/handle/1874/436876/Helmond-2019-A_plataformizac_a_o_da_web.pdf. Acesso em: 9 mar. 2026.

KHAN, Irene. Disinformation and freedom of expression. Relatório da Relatora Especial sobre liberdade de expressão (A/76/258). Nova York: ONU, 2021. Disponível em: https://docs.un.org/en/A/76/258. Acesso em: 9 mar. 2026.

KLONICK, Kate. The new governors: the people, rules, and processes governing online speech. Harvard Law Review, v. 131, n. 6, 2018. Disponível em: https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf. Acesso em: 9 mar. 2026.

KRUPIY, Tetyana. Meeting the chimera: how the CEDAW can address digital discrimination. International Human Rights Law Review, Leiden, v. 10, n. 1, p. 1–39, 2021. Disponível em: https://brill.com/view/journals/hrlr/10/1/article-

p1_1.xml?language=en&srsltid=AfmBOorkigYAsdOvdZmrh8R7p-bMayfVHpg79jnu-UFspprtZJbY7OfR. Acesso em: 9 mar. 2026.

KRUPIY, Tetyana; SCHEININ, Martin. Disability discrimination in the digital realm: how the ICRPD applies to artificial intelligence decision-making processes and helps in determining the state of international human rights law. Human Rights Law Review, Oxford, v. 23, n. 3, 2023. Disponível em: https://academic.oup.com/hrlr/article/23/3/ngad019/7237939. Acesso em: 9 mar. 2026.

LESSIG, Lawrence. Code and other laws of cyberspace. New York: Basic Books, 2000.

MANNE, Kate. Down girl: the logic of misogyny. Oxford: Oxford University Press, 2018.

MARCH, James G.; OLSEN, Johan P. The logic of appropriateness. In: GOODIN, Robert E. (ed.). The Oxford Handbook of Political Science. Oxford: Oxford University Press, 2009.

MARWICK, Alice E. Morally motivated networked harassment as normative reinforcement. Social Media + Society, v. 7, n. 2, p. 1–13, 2021. Disponível em: https://doi.org/10.1177/20563051211021. Acesso em: 9 mar. 2026.

MARWICK, Alice E.; CAPLAN, Robyn. Drinking male tears: language, the manosphere, and networked harassment. Feminist Media Studies, v. 18, n. 4, p. 543–559, 2018. Disponível em: https://doi.org/10.1080/14680777.2018.1450568. Acesso em: 9 mar. 2026.

MENDONÇA, Ricardo F., FILGUEIRAS, F., e ALMEIDA, Virgílio. "Algoritmos, desidentificação e infrapolítica da resistência." Revista Brasileira de Ciência Política, 2025. Disponível em: https://doi.org/10.1590/0103-3352.2025.44.280252. Acesso em: 9 mar. 2026.

MESECVI. Lei Modelo Interamericana para Prevenir, Punir e Erradicar a Violência Digital de Gênero contra as Mulheres. Washington: OEA, 2025. Disponível em: https://belemdopara.org/wp-content/uploads/2025/03/LEI-MODELO-INTERAMERICANA-PARA-PREVENIR-PT.docx.pdf. Acesso em: 9 mar. 2026.

MUNGER, Kevin. The YouTube apparatus. Cambridge: Cambridge University Press, 2024.

ONU. Princípios Orientadores sobre Empresas e Direitos Humanos: implementando o marco das Nações Unidas "Proteger, Respeitar e Reparar" (Princípios Ruggie). Genebra: ACNUDH, 2011.

REIDENBERG, Joel R. Lex informatica: the formulation of information policy rules through technology. Texas Law Review, v. 76, n. 3, 1997. Disponível em: https://ir.lawnet.fordham.edu/faculty_scholarship/42. Acesso em: 9 mar. 2026.

SANTINI, R. Marie et al. "Aprenda a evitar esse tipo de mulher": estratégias discursivas e monetização da misoginia no YouTube. Rio de Janeiro: NetLab UFRJ / Ministério das Mulheres, dezembro de 2024. Disponível em: https://www.gov.br/mulheres/pt-br/central-de-conteudos/publicacoes/RelatrioCompletoEstratgiasdiscursivasemonetizaodamisoginianoYouTube.pdf. Acesso em: 9 mar. 2026.

SIMONOVIC, Dubravka. Relatório da Relatora Especial sobre violência contra a mulher (A/HRC/38/47). Nova York: ONU, 2018. Disponível em: https://docs.un.org/es/A/HRC/38/47. Acesso em: 9 mar. 2026.

SOBIERAJ, Sarah. Bitch, slut, skank, cunt: patterned resistance to women's visibility in digital publics. Information, Communication & Society, v. 21, n. 11, p. 1700–1714, 2017. Disponível em: https://doi.org/10.1080/1369118X.2017.1348535. Acesso em: 9 mar. 2026.

SUZOR, Nicolas et al. Human rights by design: the responsibilities of social media platforms to address gender-based violence online. Policy & Internet, v. 11, n. 1, 2019. Disponível em: https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3.185. Acesso em: 9 mar. 2026.

ZUBOFF, Shoshana. The age of surveillance capitalism: the fight for a human future at the new frontier of power. New York: PublicAffairs, 2019.