

A CORRIDA ARMAMENTISTA LINGUÍSTICA NAS PLATAFORMAS DIGITAIS: O *ALGOSPEAK* SOB A ÓTICA DA LIBERDADE DE EXPRESSÃO

Pillar Cornelli Crestani

Doutoranda e Mestre pelo Programa de Pós-Graduação em Direito da Universidade Federal de Santa Maria (PPGD/UFSM)

Especialista em Direito Digital pela Fundação Escola Superior do Ministério Público (FMP)

Graduada em Direito pela Universidade Franciscana (UFN). Advogada

Pesquisadora do Centro de Estudos e Pesquisas em Direito e Internet da Universidade Federal de Santa Maria (CEPEDI/UFSM)

ORCID:<https://orcid.org/0000-0003-3267-4272>

e-mail: pillarcornellicrestani@gmail.com

Rafael Santos de Oliveira

Doutor em Direito pela Universidade Federal de Santa Catarina (UFSC)

Mestre em Integração Latino-Americana pela Universidade Federal de Santa Maria (UFSM)

Graduado em Direito pela Universidade Federal de Santa Maria (UFSM)

Professor no Departamento de Direito da Universidade Federal de Santa Maria (UFSM) no Mestrado e

Doutorado do Programa de Pós-graduação em Direito da UFSM (PPGD/UFSM)

ORCID:<https://orcid.org/0000-0001-5060-9779>

e-mail: advrso@gmail.com

Recebido em: 31/10/2025

Aprovado em: 11/11/2025

RESUMO

O presente artigo analisa o sistema de moderação de conteúdos das plataformas digitais e a emergência do *algospeak*, cuja questão norteadora busca responder de que forma esse fenômeno se relaciona com a liberdade de expressão dos usuários das redes sociais. Para tanto, aplicou-se o método de abordagem dedutivo, partindo da perspectiva ampla da moderação de conteúdos, atingindo a conjuntura particular do *algospeak*, combinando o método de procedimento monográfico e as técnicas de pesquisa bibliográfica e documental. A metodologia resultou na divisão do artigo em duas partes: no primeiro capítulo, elaborou-se um panorama geral sobre o sistema de moderação de conteúdos das plataformas; e, no segundo, aborda-se a relação entre *algospeak* e liberdade de expressão. Com base no estudo realizado, constatou-se que o sistema moderador é falho e opera de forma arbitrária (discricionária), com os algoritmos detendo capacidade limitada para analisar o contexto das postagens. O *algospeak*, por sua vez, emerge como uma estratégia reativa e uma ferramenta de resistência dos usuários, buscando efetivar a liberdade de expressão diante de sanções injustamente aplicadas. Todavia, o fenômeno é controverso, pois pode ser aproveitado para ocultar conteúdos indevidos, como discursos de ódio e desinformação. É imperativo, portanto, que as *big techs* aprimorem os seus sistemas de moderação, para que detenham a capacidade de analisar contextos e identificar o *algospeak* em conteúdos realmente danosos, protegendo simultaneamente a liberdade de expressão e os demais direitos fundamentais.

Palavras-chave: *algospeak*; moderação de conteúdos; liberdade de expressão.

THE LINGUISTIC ARMS RACE ON DIGITAL PLATFORMS: ALGOSPEAK FROM THE PERSPECTIVE OF FREEDOM OF EXPRESSION

ABSTRACT

This article analyzes the content moderation systems of digital platforms and the phenomenon of *algospeak*, with the aim of determining how this phenomenon relates to the freedom of expression of social media users. To achieve this, the study employed the deductive approach method, combined with the monographic procedure method and bibliographic and documentary research techniques. The methodology resulted in the division of the article into two parts: the first chapter presents an overview of content moderation systems on digital platforms, while the second addresses the relationship between *algospeak* and freedom of expression. Based on the study, it was found that content moderation systems are flawed and operate in an arbitrary (discretionary) manner, with algorithms possessing limited capacity to analyze the context of posts. *Algospeak* emerges as a reactive strategy and a tool of resistance used by users to assert their freedom of expression in the face of unjustly imposed sanctions. However, the phenomenon is controversial, as it can also be exploited to conceal inappropriate content, such as hate speech and misinformation. Therefore, it is necessary for big tech companies to improve their moderation systems so they can better interpret context and identify *algospeak* in genuinely harmful content, thus protecting freedom of expression and other fundamental rights.

Keywords: *algospeak*; content moderation; freedom of speech.

1 INTRODUÇÃO

As redes sociais não são ambientes neutros, tampouco espaços em que a liberdade de expressão pode ser exercida de forma ilimitada. A maior parte das plataformas digitais que abrigam conteúdo produzido pelos próprios usuários, como o *Facebook*, o *Instagram* e o *TikTok*, estabelece regras de conduta para seus participantes, definindo quais tipos de conteúdo não podem, em tese, ser publicados em seus domínios.

Essas normativas geralmente estão vinculadas aos termos de serviço das redes sociais ou são apresentadas em seções específicas, chamadas de “Padrões” ou “Diretrizes da Comunidade”, configurando um contrato de adesão firmado entre a plataforma e o usuário no momento em que ele cria sua conta. A presença dessas regras se justifica pela necessidade de manter o ambiente virtual minimamente saudável, prevenindo a disseminação de conteúdos inadequados, prejudiciais, enganosos ou violadores de direitos.

Em caso de violação dessas normas, são aplicadas medidas interventivas pelas plataformas, que podem variar entre a diminuição do alcance e da monetização vinculada à

conta do usuário que descumpra as diretrizes, até a remoção da postagem considerada transgressora. Por essa razão, evidencia-se que os internautas vêm se utilizando de linguagens virtuais alternativas, em formato de código, para burlar a filtragem de conteúdos das redes sociais e, assim, esquivar-se das sanções previstas pelo sistema moderador – fenômeno, esse, denominado *algospeak*.

Partindo dessa perspectiva, questiona-se: de que forma o *algospeak* se relaciona com a liberdade de expressão dos usuários das plataformas digitais? Para tanto, aplica-se o método de abordagem dedutivo, pois a pesquisa parte de uma perspectiva ampla, explicitada pelo âmbito da moderação de conteúdos das redes sociais, atingindo uma conjuntura particular, direcionando a análise da temática para a questão do *algospeak*. Também é aplicado o método de procedimento monográfico, eis que a pesquisa consiste em um estudo de caso sobre a relação entre *algospeak* e liberdade de expressão, além de se utilizar das técnicas de pesquisa bibliográfica e documental, tendo em vista que são consultadas obras de autores que dialogam com a temática proposta.

Por fim, a aplicação da referida metodologia resultou na divisão do artigo em duas partes: primeiramente, traça-se um panorama geral sobre o sistema de moderação de conteúdos das plataformas digitais, explicitando o seu contexto e operacionalização; e, em um segundo momento, aborda-se a dualidade na relação entre o fenômeno linguístico do *algospeak* e o direito à liberdade de expressão. Portanto, trata-se de uma temática ainda pouco explorada na literatura, mas que se mostra imprescindível para a compreensão das dinâmicas de poder e comunicação nas plataformas digitais contemporâneas.

2 CONSIDERAÇÕES SOBRE O SISTEMA DE MODERAÇÃO DE CONTEÚDOS DAS PLATAFORMAS DIGITAIS

Inicialmente, observa-se que a maior parte das plataformas digitais que hospedam conteúdo produzido pelos próprios usuários¹ adota sistemas de moderação próprios, com o objetivo de “sanear” o ambiente virtual, tornando-o mais seguro para os internautas e atrativo para os anunciantes, ao mesmo tempo em que busca preservar, dentro do possível, os direitos

¹ Para atender às finalidades deste artigo, convencionou-se considerar apenas a modalidade de plataforma digital de conteúdo gerado pelo usuário, que compreende “uma gama ampla de comunidades e serviços, que incluem tanto redes sociais quanto ferramentas de compartilhamento/disponibilização de vídeos, comentários etc.” (Rodrigues; Kurtz, 2020, p. 17), diferentemente de outras plataformas digitais, como as de *streaming* (*Netflix* e *Spotify*, por exemplo) e de transporte (*Uber*, por exemplo).

dos participantes e da coletividade. Nesse contexto, ressalta-se que o modelo de negócios dessas plataformas fundamenta-se na autorregulação de suas práticas de moderação de conteúdo (Rodrigues; Kurtz, 2020, p. 7).

Isso significa que as próprias redes sociais são responsáveis por regular os seus domínios, definindo parâmetros mínimos sobre os tipos de conteúdo que podem ser publicados e aqueles que, em princípio, são proibidos, além de prever a aplicação de medidas interventivas (sanções) para os casos de violações dessas diretrizes. Assim, torna-se claro que, diferentemente do que se costuma imaginar, o ambiente virtual não é neutro, tampouco isento de regras – as quais são predeterminadas pelas próprias plataformas – estando, inclusive, os usuários sujeitos à legislação estatal em caso de prática de atos ilícitos.

De acordo com Gillespie (2018, p. 21), “a moderação não é um aspecto auxiliar do que as plataformas fazem. É essencial, constitucional, definidor”. Ela consiste em “um controle prévio genérico do conteúdo que é postado por terceiros por meio de filtros” (Souza; Teffé, 2017), podendo ocorrer antes da publicação (moderação *ex ante*) e depois da publicação (moderação *ex post*) (Frazão, 2021, P. 17). Ou seja, por meio da moderação, as plataformas determinam quais as categorias de conteúdos são permitidas ou proibidas de circularem em seu meio, além de também exercerem uma curadoria do que já foi publicado em seu domínio, destacando e priorizando o alcance de determinados posts em detrimento de outros – o que acaba interferindo na “expressão propriamente dita”, pois essa dinâmica de seleção das postagens ocorre deliberadamente (Kurtz; Do Carmo; Vieira, 2021, P. 17).

Por conseguinte, verifica-se que o sistema de moderação de conteúdo das plataformas funciona de maneira dupla: combina processos automatizados com a intervenção humana. Assim, observa-se que a maioria das publicações passa por uma análise automática realizada por algoritmos e recursos de inteligência artificial, a qual pode ocorrer de forma prévia – antes de o conteúdo ser disponibilizado na plataforma – ou posterior, após sua publicação. Essa etapa prévia, em geral, atende a interesses comerciais das grandes empresas de tecnologia, como o direcionamento personalizado de conteúdos e anúncios aos usuários, com o propósito de manter o engajamento dos internautas dentro da plataforma (Kurtz; Do Carmo; Vieira, 2021, P. 17).

Quanto à análise posterior dos conteúdos, esta é realizada com o objetivo de identificar publicações que possam representar algum risco aos usuários ou que violem os termos de uso da plataforma. Essa verificação pode ocorrer de forma automatizada, por meio

dos algoritmos da própria rede social, ou ser iniciada a partir de denúncias feitas pelos usuários ao se depararem com conteúdos potencialmente infratores. A partir dessas detecções, quando aplicável, o sistema algorítmico da plataforma define as medidas a serem adotadas caso o material seja considerado inadequado aos seus padrões. O procedimento mais conhecido – e também o mais controverso, sobretudo do ponto de vista jurídico – é a remoção das postagens, embora também existam outras formas de sanção (Kurtz; Do Carmo; Vieira, 2021, P. 17).

Há que se destacar, ainda, que a atuação dos moderadores humanos, nesse contexto, é imprescindível, pois, “em geral tende a tornar a moderação mais precisa e potencialmente contextualizada”, tendo em vista a necessidade de tomada de certas decisões que a tecnologia não seria capaz de dar conta, “especialmente em zonas cinzentas para os algoritmos” (Kurtz; Do Carmo; Vieira, 2021, p. 17). Explicita-se que as postagens analisadas pelos moderadores humanos podem ter sido detectadas, de modo prévio, pelo sistema automatizado da plataforma ou, ainda, por meio das denúncias efetuadas pelos próprios usuários (conduta conhecida como “*flagging*” ou “sinalização”). Nessa perspectiva, resta evidente a necessidade de as duas modalidades de moderação de conteúdo atuarem em conjunto, em que pesem as falhas possíveis de serem cometidas por cada uma delas.

Dando continuidade ao tema, observa-se que o processamento da moderação de conteúdos é guiada pelas diretrizes ou políticas da comunidade (denominação que pode variar conforme a rede social), bem como pelos termos de uso ou termos de serviço das plataformas digitais, cuja nomenclatura também se altera de acordo com cada ambiente virtual, os quais “oferecem algumas referências sobre conteúdos restritos ou indesejáveis e práticas de moderação aplicáveis em caso de violações” (Rodrigues; Kurtz, 2020, p. 8). Em resumo, cada rede social detém as suas próprias regras, que servem para orientar a conduta dos usuários e conferir à plataforma o poder de regular o seu próprio ambiente, aplicando medidas interventivas (sanções) quando essas normas não são cumpridas.

É crucial destacar que esses termos de uso têm natureza contratual e, portanto, são obrigatórios (possuem “força de lei”). Eles constituem, na verdade, um contrato de adesão, formulado exclusivamente pela plataforma, que o usuário precisa aceitar para poder criar a sua conta e acessar a rede social (Rodrigues; Kurtz, 2020, p. 26). Ao aceitar, o internauta se sujeita integralmente às regras e às sanções estabelecidas. Embora as especificações variem ligeiramente de uma plataforma para outra, as diretrizes são, em grande parte, muito

parecidas, geralmente proibindo os mesmos tipos de conteúdo e visando a proteger os mesmos direitos, tanto dos indivíduos, quanto da coletividade.

Na sequência, destaca-se, a título de exemplo, que as diretrizes de redes sociais como *Facebook*, *Instagram* e *TikTok*, em geral, proíbem a publicação de conteúdos atrelados a todo e qualquer tipo de violência; discurso de ódio; *bullying*; nudez, pornografia infantil e exploração sexual de menores; utilização de perfis falsos, impulsionados por *bots*; desinformação. Entretanto, cada uma das plataformas dispõe sobre essas normativas de forma específica, de acordo com as peculiaridades de seus próprios mecanismos de moderação e de aplicação de medidas interventivas (Padrões..., 2025; Diretrizes..., 2025; Tiktok, 2025).

Apesar das críticas que podem ser estabelecidas em relação a esses aspectos, todavia, não se pode deixar de comentar que a proibição das referidas categorias de conteúdos – pelo menos na teoria – são importantes instrumentos de proteção à dignidade humana, buscando equilibrar o ambiente virtual, proporcionando liberdades aos usuários das plataformas, deixando-os cientes, entretanto, de que existem regras a serem seguidas, com vistas a resguardar os seus próprios direitos e mantê-los livres de potenciais danos.

Por conseguinte, verifica-se que “uma vez realizada a detecção de conteúdo potencialmente passível de moderação, cabe à plataforma avaliar se aquele conteúdo efetivamente é passível de intervenção conforme suas políticas” (Rodrigues; Kurtz, 2020, p. 68). Ou seja, se o sistema moderador constatar que o conteúdo inicialmente considerado suspeito, de fato, viola as normativas da rede social, procede-se, na sequência, à etapa da aplicação das sanções previstas. Desse modo, comenta-se que algumas dessas medidas interventivas são mais evidentes, como é o caso da remoção² de posts – e que constitui a iniciativa mais polêmica – ao passo que outras são mais ocultas, como é o caso do *shadow banning*³, e que todas elas suscitam debates acerca da transparência e da justa aplicação.

Inclusive, um dos grandes problemas do sistema de moderação de conteúdo das plataformas digitais é a ausência de aviso prévio ao usuário sobre as ações que serão tomadas em postagens suspeitas de violar as regras. Em outras palavras, as plataformas exercem o

² A remoção “consiste na eliminação completa de informação da plataforma, que passa a ser inacessível para todos os usuários” (Kurtz; Do Carmo; Vieira, 2021, p. 14). Trata-se de uma medida aplicada em casos mais graves, quando as postagens apresentam riscos potenciais de gerar algum dano ou quando, efetivamente, geram algum prejuízo ou violam algum direito dos usuários, bem como da própria coletividade.

³ Também chamada de ranqueamento, consiste no “uso de técnicas algorítmicas para determinar quais conteúdos serão priorizados em cada plataforma” (Kurtz; Do Carmo; Vieira, 2021, p. 15). O autor da postagem não é comunicado sobre a aplicação dessa restrição – apesar de ela ser perceptível, diante da diminuição do número habitual de curtidas, comentários e visualizações – constituindo, portanto, uma medida interventiva oculta (Rodrigues; Kurtz, 2020, p. 73).

direito de restringir ou remover o conteúdo de seus membros sem dar a eles a chance de revisar ou editar a publicação antecipadamente, para que ela possa ser mantida *online* (Kurtz; Do Carmo; Vieira, 2021, p. 71).

É crucial notar que todo e qualquer conteúdo está sujeito à moderação por plataformas que utilizam sistemas automatizados, mesmo que não seja, de fato, prejudicial ou infrinja as regras (Kurtz; Do Carmo; Vieira, 2021, p. 71). Isso ocorre especialmente porque os algoritmos das redes sociais são propensos a falhas, podendo detectar automaticamente palavras, frases ou imagens que apenas aparentam ser inadequadas, levando a plataforma a aplicar restrições indevidas à postagem. A dificuldade reside no fato de que a análise de conteúdo é complexa, dependendo de inúmeros fatores linguísticos e contextuais, é uma capacidade ainda muito limitada para os sistemas de Inteligência Artificial das plataformas, que se baseiam na técnica de *machine learning* (aprendizado de máquina).

Resta evidente, portanto, que o sistema de moderação das plataformas é falho e opera de forma arbitrária (discricionária), é possível, dependendo da situação, levar à violação da liberdade de expressão e de outros direitos fundamentais dos usuários. Além disso, as diretrizes das redes sociais são pouco claras e a falta de transparência sobre como a moderação é feita revela a extrema vulnerabilidade dos internautas diante das grandes empresas de tecnologia (*big techs*).

Nesse sentido, convém esclarecer que, em janeiro de 2025, a *Meta* anunciou o encerramento do programa de verificação de fatos de terceiros (“*third-party fact checking*”) no *Facebook* e *Instagram*, substituindo-o por um novo modelo denominado *Community Notes*. A proposta dessa mudança é ampliar o espaço para o discurso público, reduzindo as remoções e restrições aplicadas a temas considerados parte do “discurso *mainstream*”, e direcionando os esforços de moderação para violações de maior gravidade, como incitação à violência, abuso ou desinformação com potencial de dano real (Meta, 2025).

Além disso, a empresa alega estar intensificando o uso de tecnologias de inteligência artificial e automação na detecção e análise de conteúdos, buscando maior transparência na aplicação das normas e apresentando relatórios periódicos de execução e moderação. A *Meta* afirma que essas medidas visam a diminuir os erros de remoção indevida e aumentar a precisão das decisões, equilibrando liberdade de expressão e segurança na plataforma (Meta, 2025).

O *TikTok*, por sua vez, destaca que suas Diretrizes da Comunidade são revisadas de

forma contínua, com o objetivo de acompanhar a evolução dos comportamentos *online* e os novos tipos de risco que surgem à medida que a plataforma cresce e novas tecnologias, como a inteligência artificial, passam a integrar a criação de conteúdo. Essa atualização constante busca garantir que as políticas permaneçam eficazes e adequadas aos desafios emergentes do ambiente digital (Grover, 2025).

Ademais, a plataforma alega que vem reforçando a transparência sobre os seus processos de moderação, tornando mais claro como as regras são aplicadas em diferentes regiões e contextos culturais. Essa abordagem pretende assegurar que a aplicação das normas seja coerente, compreensível e sensível às particularidades locais, fortalecendo a confiança dos usuários e promovendo um ambiente mais seguro e equilibrado para a expressão criativa (Grover, 2025).

O *TikTok* também divulgou o Relatório de Aplicação das Diretrizes da Comunidade referente ao 1º trimestre de 2025, apresentando as ações proativas que teriam sido adotadas pela plataforma para identificar e remover conteúdos que violam suas Diretrizes da Comunidade, buscando preservar a integridade do espaço de convivência e interação global. Segundo o relatório, teriam sido removidos 211 milhões de vídeos em todo o mundo durante o período, e, desse total, 184.378.987 vídeos teriam sido identificados e excluídos por meio de tecnologias de detecção automatizada, enquanto 7.525.184 vídeos teriam sido posteriormente restaurados após revisão adicional. A taxa global de remoção proativa teria alcançado 99,0%, com 94,3% do conteúdo sinalizado que foi removido em até 24 horas após a sua publicação inicial (TBS News, 2025).

Portanto, o ideal de transparência nas redes sociais exige que as plataformas ajam de forma semelhante ao devido processo legal, começando por garantir que os usuários tenham total clareza sobre as suas políticas e termos de uso – o que inclui a divulgação detalhada dos métodos de filtragem e análise de conteúdo, bem como dos critérios de restrição específicos utilizados para eliminar omissões, contradições e ambiguidades nas postagens “irregulares”. Além disso, é fundamental que o usuário seja notificado sobre a infração para ter a chance de editar o conteúdo antes da ação da plataforma, assegurando-lhe o direito de contestar a medida sofrida, garantindo que o sistema de moderação possa revisar o caso e, se necessário, reverter a restrição aplicada (Kurtz; Do Carmo; Vieira, 2021, p. 20).

Inclusive, não se pode deixar de ressaltar o fato de as redes sociais adotarem o regime da “economia de atenção”, pelo qual o maior interesse das plataformas reside na busca

pelo maior tempo de permanência dos usuários em seu domínio, para consumir conteúdos e gerar engajamento às postagens (Wu, 2022). Percebe-se, assim, que as *big techs* escondem as suas verdadeiras intenções, agindo de forma contrária às suas próprias regras. As diretrizes e políticas, na prática, parecem servir apenas como um instrumento de fachada para que as empresas mostrem ao público – e principalmente aos seus anunciantes – que estão monitorando e regulando seus ambientes. O objetivo final, portanto, é criar a impressão de que o espaço virtual não está abandonado à própria sorte, mas sim sob controle.

Contrariando o que afirmam, as diretrizes das redes sociais parecem servir aos seus próprios interesses, e não à proteção dos direitos dos usuários e da coletividade. Em suma, é como se as plataformas fossem um "lobo em pele de cordeiro", e que toda essa falta de transparência, assim como as contradições e falhas do sistema (sejam elas acidentais, devido a desajustes tecnológicos, ou intencionais, para concretizar os seus objetivos), causam um impacto direto nos direitos fundamentais dos usuários, afetando sua dignidade, seus direitos personalíssimos, o direito à informação e, especialmente, a liberdade de expressão.

Nessa perspectiva, resta evidente que as *big techs* recebem inúmeras recomendações para que aprimorem seus sistemas de moderação e, assim, evitem a violação dos direitos dos usuários. Paralelamente, há uma pressão cada vez maior para a regulamentação dessas plataformas digitais, uma vez que o objetivo é que elas sejam responsabilizadas civilmente por manter conteúdo prejudicial em seus domínios, devido às falhas (propositais ou não) em seus mecanismos de moderação.

No próximo capítulo, será abordado, à luz do direito à liberdade de expressão, o fenômeno do *algospeak* – mediante o qual os usuários das redes sociais tentam driblar a moderação de conteúdos, a fim de tentarem escapar da aplicação de medidas interventivas pelas plataformas digitais.

3 O FENÔMENO DO *ALGOSPEAK* SOB A ÓTICA DA LIBERDADE DE EXPRESSÃO DOS USUÁRIOS DAS PLATAFORMAS DIGITAIS

De acordo com o que restou exposto no capítulo anterior, evidencia-se que o sistema de moderação de conteúdo das plataformas digitais opera de maneira dupla: combina processos automatizados com a intervenção humana, e que a maioria das publicações passa por uma análise automática realizada por algoritmos e recursos de inteligência artificial,

podendo ocorrer de forma prévia (antes de o conteúdo ser disponibilizado na rede) ou posterior (após a sua publicação).

Entretanto, evidencia-se que o sistema de moderação automatizado apresenta falhas, uma vez que a simples utilização de termos linguísticos associados a temas considerados proibidos pelas plataformas costuma ser suficiente para que a moderação automática identifique o conteúdo como violador de seus termos de uso e, em razão disso, aplique medidas restritivas. Nessa perspectiva, reitera-se que a mais recorrente delas é o *shadow banning*, que resulta na diminuição do alcance das publicações e na redução dos índices de engajamento e monetização do perfil do usuário considerado “infrator” (Crestani, 2022).

Por conseguinte, ressalta-se que o funcionamento desse tipo de mecanismo varia conforme a plataforma, eis que cada uma delas detém o seu próprio sistema de moderação de conteúdo. Ainda assim, a título ilustrativo, pode-se mencionar o caso do *Instagram*, onde a utilização de palavras, faladas ou escritas, que remetem a temas vedados pela rede social, como “morte”, “abuso” ou “ódio”, costuma provocar uma espécie de “boicote” ao conteúdo. Nesses casos, os algoritmos da plataforma passam a limitar significativamente o alcance da publicação classificada como inadequada ou sensível (Crestani, 2022).

Em razão disso, é comum que criadores de conteúdo relatem queda no engajamento de seus perfis ou de seus *posts* devido ao uso de termos considerados proibidos pelas plataformas – evidenciando as falhas do sistema de moderação automatizado, que não consegue compreender o contexto das publicações. Assim, a mera menção a termos sensíveis não deve ser interpretada como violação dos termos de uso, já que conteúdos informativos, como notícias sobre crimes ou mortes, não configuram apologia à violência, tornando injustificadas as medidas restritivas aplicadas nesses casos (Crestani, 2022).

Diante disso, para evitar a redução do engajamento de seus perfis e da monetização de suas publicações, muitos usuários passam a adotar estratégias para contornar os mecanismos de moderação automatizada das plataformas – como é o caso do *algospeak*, cujo termo “reflete sua origem etimológica, combinando “algo-” (de “algoritmo”) com “speak” (linguagem), denotando uma linguagem cifrada para evadir a censura algorítmica” (Chu, 2025, p. 8). No que tange ao conceito desse fenômeno, tem-se o seguinte:

Algospeak é uma linguagem que compreende um conjunto de códigos e símbolos para contornar algoritmos e filtros de moderação de conteúdo nas redes sociais. Os usuários modificam ou substituem, por meio de uma comunicação criativa, termos considerados “impróprios” ou “sensíveis” pelas plataformas, empregando expressões

alternativas ou símbolos que mantêm o significado original, mas dificultam a detecção pelos sistemas automatizados - a exemplo de *mat4r* (matar), *est3pro* (estupro), *ass3dio* (assédio) (Barreto et al., 2025, p. 194).



Destaca-se que, embora o *algspeak* seja um fenômeno recente, ele não surge de forma inédita: ao longo do tempo, outras adaptações linguísticas também foram criadas em contextos de restrição comunicativa, como o *Leetspeak*, popularizado na década de 1980, que trocava letras por números para formar uma escrita codificada; e o *Voldemorting*⁴, em que se evitava pronunciar determinadas palavras para escapar de mecanismos de rastreamento. Esses exemplos, por sua vez, evidenciam a capacidade humana de desenvolver alternativas linguísticas para contornar limitações impostas à comunicação (Chu, 2025, p. 10).

Nessa perspectiva, também não se pode deixar de mencionar que o *algspeak* passou a ter notoriedade na pandemia da Covid-19, à medida em que as plataformas reduziram o alcance de conteúdos relacionados à crise sanitária para evitar a propagação de desinformação sobre o assunto. Por essa razão, os usuários das redes sociais passaram a se referir à pandemia como “turnê de reunião dos Backstreet Boys”; “panini” ou “panda express”, como forma de driblar a moderação automática, evitando a aplicação de medidas interventivas pelas plataformas (Lorenz, 2022).

O fato é que a criatividade e a diversidade do *algspeak* manifestam-se em múltiplas estratégias de codificação que surgem constantemente em diferentes plataformas e idiomas. Tais variações refletem a inventividade dos usuários ao adaptar sua linguagem para driblar os filtros automatizados das plataformas (Chu, 2025, p. 10). A seguir, apresentam-se alguns exemplos que demonstram os distintos métodos por meio dos quais os usuários codificam suas postagens:

<i>Algspeak</i>	
Substituição de letras por números e outros caracteres semelhantes	“morte” por “m0r73”; “abuso” por “@bu\$0”; “ódio” por “*d10” (Crestani, 2022).
Eufemismos em substituição a termos proibidos	“Perdi a vontade de jogar”, para se referir a uma ideação suicida (Chu, 2025, p. 10).

⁴ Expressão inspirada no personagem Lord Voldemort da série Harry Potter, que era referido como “aquele que não deve ser nomeado”, evitando-se citar seu nome, por medo ou superstição (CHU, 2025, p. 10).

Utilização de emojis para se referir a um assunto ou termo específico	 (Berinjela)  (Pêssego) Por lembrarem vagamente partes da anatomia humana, foram sendo utilizados para se referir a conteúdos sexuais. Entretanto, em 2019, tanto o <i>Facebook</i> quanto o <i>Instagram</i> tomaram medidas para bloquear o seu uso com conotação sexual (Kreuz, 2023).
---	---

Fonte: Autores.

Diante disso, evidencia-se que os internautas vão readaptando a sua comunicação, à medida em que os algoritmos das plataformas vão se aperfeiçoando e detectando os códigos criados pelos usuários para burlar a filtragem do sistema de moderação (Barreto *et al.*, 2025, p. 200), gerando uma “corrida armamentista linguística” (Kreuz, 2023). Portanto, o *algospeak* marca uma nova fase no desenvolvimento da linguagem, em que a comunicação virtual é contornada pela interação entre os humanos e a tecnologia (Chu, 2025, p. 11). Nesse sentido:




O surgimento do Algospeak representa um fenômeno linguístico paradigmático na comunicação digital, mostrando como a moderação algorítmica por IA paradoxalmente estimula a evolução da escrita. Este artifício linguístico emerge como uma estratégia reativa dos usuários para subverter filtros automatizados e sistemas de moderação, demonstrando a capacidade adaptativa da linguagem às novas realidades sociotecnológicas (Chu, 2025, p. 7).

Partindo dessas considerações, constata-se que o *algospeak* constitui um fenômeno controverso, especialmente, em razão da falta de transparência e das falhas cometidas pelo sistema de moderação de conteúdos das plataformas, cujos algoritmos não detêm a capacidade de analisar e compreender variações linguísticas, bem como o contexto das postagens dos usuários. A partir disso, entende-se que essa tentativa de burlar os mecanismos de moderação das redes sociais, por meio do *algospeak*, também pode ser aproveitada por usuários que, de fato, publiquem conteúdos proibidos pela plataforma – como discursos de ódio, por exemplo – tornando mais difícil a identificação de publicações que realmente violam os termos de serviço, com a consequente aplicação das sanções cabíveis (Crestani, 2022).

Esse é o caso ilustrado na série “Adolescência”, que estreou na *Netflix* no ano de 2025, em que o protagonista é adepto da “Cultura Incel”⁵, cuja comunidade virtual interage

⁵ O termo “Incel” é abreviação de “involuntary celibate” (celibatário involuntário), que se refere a homens que

por meio de *emojis*, cada um contendo um significado específico, conforme a tabela abaixo, baseada nas informações obtidas em Chu (2025, p. 5):

<i>Algospeak</i> – Cultura Incel	
	Remete ao movimento “Red Pill”, ligado a ideologias misóginas e extremistas.
	Tesão
	Estou interessada, mas não em sexo.

Fonte: Autores.

Portanto, verifica-se que a Cultura Incel está atrelada ao machismo, à misoginia e ao extremismo, propagando discursos de ódio, bem como outras categorias que violam as políticas das plataformas – como é o caso de conteúdos de teor sexual, mencionados anteriormente. Nesse sentido, não restam dúvidas de que as postagens dos usuários dessa comunidade deveriam sofrer a aplicação de medidas interventivas, especialmente, pelo fato de violarem direitos. Contudo, isso acaba por ser dificultado pela utilização do *algospeak* – o qual “se configura como uma adaptação estratégica fundamental para evitar que conteúdos, palavras ou discussões sejam bloqueados, removidos ou penalizados” (Chu, 2025, p. 7-8).

Por outro lado, verifica-se que usuários cujos perfis nas plataformas são dedicados a conteúdos educacionais ou jornalísticos, por vezes, acabam penalizados injustamente pela aplicação de medidas interventivas. Isso porque a mera menção a uma palavra associada a um tema proibido em determinada rede social não implica que todo o conteúdo seja inadequado. Por exemplo, se uma conta no *Instagram* divulgar a ocorrência de um crime de homicídio, não significa que o perfil esteja incentivando a violência. Assim, não há fundamento legítimo para a plataforma adotar medidas restritivas nesse caso, uma vez que o conteúdo, claramente, não configura a violação dos termos de serviço, tampouco se enquadra como publicação imprópria (Crestani, 2022). Porém, conforme já explicitado anteriormente, os algoritmos da moderação automática das redes sociais não detêm a capacidade de analisar o contexto das

afirmam não conseguir estabelecer relacionamentos amorosos ou sexuais, apesar de sua pretensão em tê-los. Em comunidades virtuais, esses indivíduos costumam expressar frustrações relacionadas à rejeição e ao sentimento de exclusão social, frequentemente sustentadas por uma visão de mundo que julga injustas as normas que favorecem homens considerados atraentes ou bem-sucedidos. Essa subcultura adota conceitos como “Hipergamia” (a crença de que mulheres buscam parceiros de status mais elevado) e “Red Pill” (em referência ao filme “Matrix”, associada à ideia de “despertar” para uma suposta consciência sobre a realidade distorcida das dinâmicas de gênero. Em alguns casos, essas crenças assumem contornos radicais e misóginos, levando determinados membros a responsabilizar a sociedade e as mulheres por suas frustrações, o que, em situações extremas, tem resultado em episódios de violência motivados por ódio (Pinto, 2024, p. 8).

postagens, para verificar se determinado conteúdo viola, de fato, as diretrizes da plataforma, razão pela qual a simples menção a um termo considerado indevido já é alvo da aplicação de sanções – como a redução do alcance do *post*. Em função disso, inúmeros criadores de conteúdo das redes sociais acabam se valendo do *algospeak* como alternativa às restrições impostas indevidamente pelas plataformas.

Nessa perspectiva, destaca-se o estudo formulado por Steen, Yurechko e Klug (2023), que analisa as estratégias de criadores de conteúdo pertencentes a grupos minoritários diante da moderação algorítmica do *TikTok*. A pesquisa revela que esses usuários recorrem ao *algospeak* (por meio do uso de termos alternativos, erros propositais e substituições de letras) para evitar remoções injustas e garantir a visibilidade de temas relacionados a identidades marginalizadas, como questões raciais⁶, de gênero, saúde mental e sexualidade. O estudo conclui que o *TikTok* opera em uma lógica paradoxal: embora dependa da criatividade de grupos minoritários, tende a marginalizá-los por meio de filtros e bloqueios algorítmicos. Assim, o *algospeak* torna-se uma ferramenta de resistência e adaptação, permitindo que esses criadores continuem a discutir temas sensíveis e socialmente relevantes dentro dos limites impostos pela plataforma (Steen; Yurechko; Klug, 2023).

Diante de tudo o que foi exposto até o momento, depreende-se que o fenômeno linguístico do *algospeak* está diretamente relacionado ao direito à liberdade de expressão dos usuários das redes sociais, uma vez que a aplicação injusta de medidas interventivas, pelas plataformas, em relação a conteúdos que não violam as suas diretrizes, conforme os casos mencionados anteriormente, pode ser considerada uma forma de censura. Em contrapartida, não se pode falar em violação da liberdade de expressão dos usuários que se valem do *algospeak* para driblar a moderação automática das plataformas nas situações em que, de fato, os discursos são considerados inapropriados, sensíveis e ofensivos aos direitos de terceiros, como ocorre nos exemplos citados, relacionados à propagação de desinformação sobre a pandemia da Covid-19 e à Comunidade Incel.

Isso porque a internet é um ambiente que exige o mínimo de regulamentação, e não existe um antagonismo entre a regulação exercida pelas plataformas digitais e o direito à liberdade de expressão, pois a primeira é fundamental para assegurar a efetividade da segunda, “evitando as inúmeras distorções e manipulações que têm ocorrido no ambiente

⁶ Criadores de conteúdo negros, em especial, relatam que palavras como “racismo”, “negro”, “preta” e “nazismo” reduzem o alcance de seus vídeos, o que os leva a desenvolver táticas de subversão, como alterar grafias ou suavizar discursos para escapar da penalização automatizada (Jacques, 2025, p. 90-91).

virtual” (Frazão, 2021, p. 78). Sob essa ótica, é importante reconhecer que a liberdade de expressão constitui um direito humano e fundamental que garante o livre exercício da “manifestação da atividade intelectual, artística, científica e de comunicação, independentemente de censura ou licença” (Brasil, 1988).

Conforme observa Ledesma (2004, p. 1-17), não há um conceito único que defina plenamente esse direito, o qual se caracteriza por múltiplos paradoxos. De todo modo, trata-se de um pilar essencial da sociedade democrática e, embora constitua “condição indispensável de quase todas as outras liberdades”, não é um direito absoluto, exigindo que seu exercício seja harmonizado com a proteção dos demais direitos fundamentais. Diante disso, entende-se que as plataformas digitais detêm legitimidade para aplicar restrições a perfis e postagens que propagam discursos de ódio, misoginia, desinformação – e tantas outras categorias de conteúdos sensíveis e danosos – tendo em vista que a dignidade humana, o direito à informação e o direito à saúde, a título de exemplo, são prioritários em relação à liberdade de expressão dos usuários das redes sociais.

Por outro lado, há que se reconhecer a imprescindibilidade de as *big techs* aperfeiçoarem os seus sistemas de moderação, a fim de que os algoritmos detenham a capacidade de analisar contextos linguísticos, sociais e culturais das postagens, para poder aplicar as suas medidas interventivas com mais segurança e acurácia, evitando o cometimento de equívocos e de discricionariedades – de modo que os internautas não precisem se valer do *algospeak* para escapar das sanções aplicadas pelas plataformas. Igualmente, é necessário que a moderação automática seja capaz de identificar as estratégias de ocultação de termos realmente problemáticos, por meio da utilização do *algospeak*, sobretudo quando o conteúdo for, de fato, sensível, enganoso ou violador de direitos de terceiros.

Nessa perspectiva, não se pode deixar de comentar sobre a capacidade de um Large Language Model (LLM), especificamente o GPT-4, para decifrar esses termos codificados. Um estudo desenvolvido por Fillies e Paschke (2024) demonstra que o GPT-4 consegue traduzir com alta precisão (79,4%) os termos de *algospeak* sem contexto e atinge uma precisão de 98,5% quando é fornecida uma frase de exemplo. As descobertas sugerem que os LLMs representam uma solução promissora para aprimorar os futuros sistemas de moderação de conteúdo, fornecendo *datasets* “limpos” ao decodificar a linguagem evasiva. A pesquisa conclui que a aplicação direta de um grande modelo de linguagem pode decifrar o *algospeak* avançado e, no futuro, permitir que ele seja incluído em modelos de detecção de conteúdo

mais padronizado (Fillies; Paschke, 2024).

Portanto, entende-se que a solução para os desafios relacionados à moderação de conteúdos não é simples, exigindo o comprometimento efetivo das plataformas com a transparência e a proteção simultânea da liberdade de expressão e das demais garantias dos usuários, buscando harmonizá-los sempre que possível. É nesse ponto de intersecção, entre o avanço tecnológico contextual e a transparência regulatória, que reside a complexa busca pelo equilíbrio dos direitos no ambiente digital.

4 CONCLUSÃO

A partir de tudo o que foi discutido anteriormente, evidenciou-se que o sistema de moderação de conteúdos das plataformas digitais, embora se apresente como um mecanismo destinado a promover segurança, integridade e equilíbrio no ambiente virtual, revela-se permeado por contradições e falhas estruturais. Fundamentado em um modelo de autorregulação – uma prática que difere da censura ou moderação exercida diretamente pelo Estado – esse sistema opera sob lógicas essencialmente econômicas, priorizando a manutenção do engajamento e a atratividade para anunciantes em detrimento da efetiva proteção dos direitos dos usuários.

Verificou-se, ainda, que os processos de moderação carecem de transparência, proporcionalidade e coerência: os algoritmos, ainda limitados em sua capacidade de interpretação contextual, frequentemente classificam como inapropriados conteúdos legítimos, aplicando sanções sem aviso prévio ou possibilidade de revisão – o que reforça a assimetria de poder entre internautas e plataformas, acentuando a vulnerabilidade destes diante das decisões unilaterais das *big techs*.

Nessa perspectiva, o *algospeak* surge como uma estratégia criativa de resistência, permitindo aos usuários expressarem-se e discutirem temas relevantes sem sofrer a aplicação indevida de medidas restritivas pelas redes sociais. Portanto, trata-se de um fenômeno complexo, que, além de revelar as fragilidades e contradições dos sistemas automatizados de moderação de conteúdo nas plataformas digitais, também gera uma verdadeira “corrida armamentista linguística” por parte dos internautas.

A partir disso, a pesquisa demonstrou que a detecção meramente lexical – baseada em palavras associadas a temas sensíveis – conduz a restrições injustificadas, penalizando

usuários que não violam os termos de uso, especialmente aqueles que produzem conteúdo educativo, informativo ou jornalístico. Entretanto, o fenômeno também apresenta um caráter ambíguo: enquanto constitui ferramenta legítima de defesa da liberdade de expressão, pode, igualmente, ser utilizado para ocultar discursos de ódio, desinformação e outras formas de violação de direitos. Assim, a relação entre *algospeak* e liberdade de expressão é direta: a aplicação injusta de medidas interventivas em conteúdos que não violam as diretrizes pode ser interpretada como uma forma de censura. Por outro lado, não se pode alegar violação da liberdade de expressão quando o *algospeak* é utilizado para burlar a moderação de discursos inapropriados, ofensivos ou que violam direitos de terceiros, como a propagação de desinformação ou de discursos de ódio, por exemplo.

Em face dessas considerações, por fim, conclui-se que a solução para os desafios da moderação de conteúdo é complexa, o que torna necessário que as *big techs* aprimorem seus sistemas de moderação para que os algoritmos sejam capazes de analisar contextos linguísticos, sociais e culturais com maior acurácia e segurança. Igualmente, a moderação automática deve evoluir para identificar o *algospeak* quando utilizado para ocultar conteúdo que, de fato, viole as políticas e os direitos de terceiros. Assim, de modo a evitar a vulnerabilidade de seus usuários, é crucial que as plataformas demonstrem um comprometimento efetivo com a transparência, garantindo que os internautas tenham clareza sobre as políticas de filtragem; sejam notificados previamente sobre infrações e tenham o direito de contestar as medidas aplicadas. Em suma, o desafio reside em equilibrar a autorregulação exercida pelas plataformas com a proteção da liberdade de expressão e dos demais direitos fundamentais dos usuários das redes e da coletividade.

REFERÊNCIAS

BARRETO, Matheus Silva *et al.* Algospeak na ordem discursiva do Instagram: a subversão do algoritmo. **Boletim de Conjuntura (BOCA)**, Boa Vista, v. 21, n. 62, p. 187–204, 2025. DOI: <https://doi.org/10.5281/zenodo.14941940>. Disponível em: <https://revista.ioles.com.br/boca/index.php/revista/article/view/6640>. Acesso em: 10 out. 2025.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 10 out. 2025.

CHU, Leandro Alves. O impacto da IA na escrita na era das tendências da internet. **SciELO Preprints**, 2025. DOI: <https://doi.org/10.1590/SciELOPreprints.12995>. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/12995/version/13667>. Acesso em: 10 out. 2025.

CRESTANI, Pillar Cornelli. “Algospeak”: driblando a moderação de conteúdos nas redes sociais. **Centro de Estudos e Pesquisas em Direito e Internet da UFSM**, Santa Maria, 4 maio 2022. Disponível em: <https://cepediufsm.wordpress.com/2022/05/04/algospeak-driblando-a-moderacao-de-conteudos-nas-redes-sociais/>. Acesso em: 10 out. 2025.

DIRETRIZES da Comunidade do YouTube. **YouTube**. 2025. Disponível em: <https://support.google.com/youtube/answer/9288567?hl=pt-BR>. Acesso em: 10 out. 2025.

FILLIES, Jan; PASCHKE, Adrian. Simple LLM based Approach to Counter Algospeak. In: WORKSHOP ON ONLINE ABUSE AND HARMS (WOAH), 8., 2024, Cidade do México. **Proceedings...** Cidade do México: Association for Computational Linguistics, 2024. p. 136-145.

FRAZÃO, Ana. **Dever geral de cuidado das plataformas diante de crianças e adolescentes**. São Paulo: Instituto Alana, 2021. 261 p. Disponível em: <https://criancaconsumo.org.br/wp-content/uploads/2021/11/dever-geral-de-cuidado-dsplataformas.pdf>. Acesso em: 15 jan. 2022.

GILLESPIE, Tarleton. **Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media**. New Haven; London: Yale University Press, 2018. Disponível em: https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media/link/5dfcffa3a6f6c2837318e10/download. Acesso em: 20 out. 2021.

GROVER, Sandeep. **Create, discover, and connect on TikTok with simpler and stronger Community Guidelines**. TikTok, 15 ago. 2025. Disponível em: <https://newsroom.tiktok.com/community-guidelines-update-newsroom?lang=pt-BR>. Acesso em: 11 nov. 2025.

JACQUES, Janderson Pereira. **Onde estão as pessoas negras na for you?: racismo algorítmico e a disputa por visibilidade no Tiktok**. 2025. 126 f. Dissertação (Mestrado em Comunicação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2025. Disponível em: <https://lume.ufrgs.br/handle/10183/292475>. Acesso em: 10 out. 2025.

KREUZ, Roger. What is ‘algospeak’? Inside the newest version of linguistic subterfuge. **The Conversation**, 13 abr. 2023. Disponível em: <https://theconversation.com/what-isalgorithmspeak-inside-the-newest-version-of-linguistic-subterfuge-203460>. Acesso em: 10 out. 2025.

KURTZ, Lahis Pasquali; DO CARMO, Paloma Rocillo Rolim; VIEIRA, Victor Barbieri Rodrigues. **Transparência na moderação de conteúdo: tendências regulatórias nacionais**.

Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2021. *E-book*. Disponível em: <https://bit.ly/3xjAUka>. Acesso em: 06 jul. 2021.

LEDESMA, Héctor Faúndez. **Los límites de la libertad de expresión**. México: Instituto de Investigaciones Jurídicas de la Universidad Nacional Autónoma de México, 2004.

LORENZ, Taylor. Internet ‘algospeak’ is changing our language in real time, from ‘nip nops’ to ‘le dollar bean’. **The Washington Post**, 08 abr. 2022. Disponível em: <https://www.washingtonpost.com/technology/2022/04/08/algospeak-tiktok-le-dollar-bean/>. Acesso em: 10 out. 2025.

META. **More speech and fewer mistakes**. Menlo Park: Meta Newsroom, 7 jan. 2025. Disponível em: https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/?utm_source=https://www.google.com/search?q=chatgpt.com. Acesso em: 10 nov. 2025.

PADRÕES da Comunidade do Facebook. **Meta**. 2025. Disponível em: <https://transparency.meta.com/pt-br/policias/community-standards/>. Acesso em: 10 out. 2025.

PINTO, Júlio César Aranha Serra. **O movimento da supremacia masculina e propagação de ódio nas redes: um ensaio sobre a cultura incel**. 2024. 66 f. TCC (Bacharelado em Direito) – Centro Universitário Unidade de Ensino Superior Dom Bosco, São Luís, 2024. Disponível em: <http://repositorio.undb.edu.br/bitstream/areas/1270/1/J%c3%9aLIO%20C%c3%89SAR%20ARANHA%20SERRA%20PINTO.pdf>. Acesso em: 10 out. 2025.

RODRIGUES, Gustavo; KURTZ, Lahis. **Transparência sobre moderação de conteúdo em políticas de comunidade**. Belo Horizonte: Instituto de Referência em Internet e Sociedade, 2020. Disponível em: <https://irisbh.com.br/wp-content/uploads/2020/10/Transpar%3%AAncia-sobre-modera%3%A7%C3%A3o-deconte%3%BAado-em-pol%3%ADticas-de-comunidade-IRIS.pdf>. Acesso em: 20 out. 2020.

SOUZA, Carlos Affonso; TEFFÉ, Chiara Spadaccini de. Responsabilidade dos provedores por conteúdos de terceiros na internet. **Conjur**. 23 jan. 2017. Disponível em: https://itsrio.org/wp-content/uploads/2017/01/ConJur-Responsabilidade-do-provedor-porconteu_dode-terceiro-na-internet.pdf. Acesso em: 25 jul. 2020.

STEEN, Ella; YURECHKO, Kathryn; KLUG, Daniel. You can (not) say what you want: Using algospeak to contest and evade algorithmic content moderation on TikTok. **Social Media+ Society**, v. 9, n. 3, p. 20563051231194586, 2023. Disponível em: <https://journals.sagepub.com/doi/full/10.1177/20563051231194586>. Acesso em: 10 out. 2025.

TBS NEWS. TikTok releases Q1 2025 Community Guidelines Enforcement Report. 23 jul. 2025. Disponível em: https://www.tbsnews.net/economy/corporates/tiktok-releases-q1-2025-community-guidelines-1196201?utm_source=chatgpt.com. Acesso em: 10 nov. 2025.

TIKTOK. Diretrizes da comunidade: temas sensíveis e para adultos. 14 ago. 2025.
Disponível em: <https://www.tiktok.com/community-guidelines/pt/sensitive-mature-themes>.
Acesso em: 10 out. 2025.

WU, Tim. **A Proposal for Network Neutrality.** Charlottesville: University of Virginia, 2002.
Disponível em: <http://www.timwu.org/OriginalNNProposal.pdf>. Acesso em: 15 ago. 2025.