

# A MODERAÇÃO DE CONTEÚDO: o problema do discurso de ódio e sua vagueza

## CONTENT MODERATION: the problem of hate speech and its vague meaning

Lucila Gabriella Maciel Carneiro VILHENA<sup>1</sup>  
Universidade Estadual da Paraíba | Brasil  
Universidade de Lisboa | Portugal

### Resumo

Este artigo visa discutir o papel das grandes empresas de comunicação da Internet em relação à moderação de conteúdo odioso. Também se pretende analisar o problema que a vagueza do conceito de discurso de ódio traz para as empresas. É dizer, sua falta de significado, além de dificultar o reconhecimento deste discurso, também traz insegurança quanto à necessidade de sua exclusão do debate público, sobretudo quando esta exclusão se pretenda necessária para proteger a própria sociedade e manter parâmetros de segurança no ambiente digital. Para tal, discutiremos alguns aspectos acerca da normativa europeia, com o fim de apontá-la como referência para a condução de toda problemática aqui estudada, típica da atualidade.

### Palavras-chave

Direitos fundamentais; discurso de ódio; moderação de conteúdo;

### Abstract

This article aims to discuss the role of major internet media companies in relation to the moderation of hateful content. It is also intended to analyze the problem that the vagueness of the concept of hate speech brings to companies. That is to say, its lack of meaning, in addition to making it difficult to recognize this discourse, also brings uncertainty about the need for its exclusion from public debate, especially when this exclusion is intended to be necessary to protect society itself and maintain security parameters in the digital environment. To this end, we will discuss some aspects of the European regulations, in order to point them out as a reference for the conduct of all the problems studied here, typical of today.

### Keywords

Fundamental rights; hate speech; Content moderation.

EIXO TEMÁTICO

RECEBIDO EM 11 DE NOVEMBRO DE 2023  
ACEITO EM 19 DE ABRIL DE 2023

<sup>1</sup> Professora da Universidade Estadual da Paraíba. Contato: lucilavilhena@hotmail.com.

## Introdução

Nos dias atuais, ao passo em que há grande facilidade e rapidez na troca de informações e de opiniões na Internet, há também grande dificuldade em obter a justa responsabilização de pessoas que abusam dos direitos relativos à liberdade de informação e de expressão, sobretudo devido à própria natureza do ambiente virtual onde as pessoas se colocam distantes, “protegidas” pela tela do computador e até mesmo, em muitos casos, pelo anonimato.

Neste sentido, conforme percebemos, as autoridades públicas, diante de sua incapacidade para solucionar um problema tão específico, novo e fluido, têm passado para as mãos das empresas privadas a grande responsabilidade de promover um espaço democrático e saudável de debate público *on-line*, o que acaba por resultar, também, em grande poder concentrado nestas entidades.

Como sabemos, essas referidas empresas são hoje mais poderosas do que muitos Governos e países, elas têm capacidade de moldar a opinião pública de uma forma que pode trazer consequências para toda a humanidade.

Com efeito, em decorrência da liberdade exercida sob o ponto de vista da informação, bem como, dos efeitos colaterais do avanço tecnológico, essas plataformas também têm sofrido cada vez mais pressão – tanto por parte das autoridades públicas, quanto da sociedade em geral - para eliminar conteúdo falso e/ou ofensivo de seu ambiente, e tal pressão, como é previsto, nem sempre vem munida de boas intenções. Muitos são os relatos sobre tentativas coordenadas de manipulação da opinião pública, que ocorrem, principalmente, através do uso de contas e usuários falsos nas redes sociais.

Diante do exposto, a discussão que ora propomos é a de saber qual é o papel das grandes empresas de comunicação na Internet, sobretudo as mais relevantes, como *Google, Amazon, Facebook, Apple e Microsoft*, quanto à regulação de temas sensíveis, a exemplo da liberdade de expressão;

nomeadamente se, em se tratando de ambiente privado, estas empresas têm o dever de pautar suas políticas de moderação com base nos princípios internacionais dos Direitos Humanos; e por fim, se a falta de clareza do conceito de discurso de ódio impossibilita a ponderação dos direitos em colisão e, conseqüentemente, a própria justiça.

## **A vanguarda da normativa europeia**

Em 2016, a Comissão Europeia apresentou um código de conduta juntamente com algumas empresas de tecnologia da informação, nomeadamente: *Facebook, Microsoft, Twitter e Youtube*, chamado de “Código de conduta para combater o discurso de ódio ilegal *on-line*”, cujo objetivo foi “promover e facilitar a liberdade de expressão no mundo *on-line*” (COMISSÃO EUROPEIA, 2016).

Neste referido documento, as empresas de Tecnologia da Informação compartilham o compromisso da Comissão Europeia e dos Estados Membros da União Europeia em combater o discurso de ódio ilegal *on-line*. Discurso de ódio aqui entendido em conformidade com a Decisão-Quadro 2008/913/JAI, de 28 de novembro de 2008, relativa à luta contra certas formas e manifestações de racismo e xenofobia, em cujo texto encontramos a seguinte definição: Discurso de ódio é a “incitação pública à violência ou ao ódio contra um grupo de pessoas ou os seus membros, definido por referência à raça, cor, religião, ascendência ou origem nacional ou étnica”. (CONSELHO, 2008).

Na elaboração daquele código de conduta, as empresas se comprometeram a observar as regras e diretrizes europeias que proíbem o incitamento à violência e à conduta odiosa. Com efeito, as empresas também assumiram o compromisso de, dentre outros atos, manter processos claros e eficazes para a análise das notificações referentes a discursos de ódio que lhes fossem apresentadas, bem como, promover a rápida remoção (em até 24 horas) de tais conteúdos, sempre analisando-os à luz das suas políticas internas,

obviamente, mas também à luz da legislação europeia e das leis nacionais que transpõem aquela Decisão-Quadro.

Conforme notamos, a justificativa para a adoção de tais medidas pelo referido código de conduta para combater o discurso de ódio ilegal *on-line*, foi precisamente a necessidade de defender o direito à liberdade de expressão, embora nem as autoridades europeias e nem as empresas expliquem exatamente como isso se daria.

O referido documento ainda previu a avaliação periódica, levada a cabo por organizações com sede nos Estados-membros, do comportamento das empresas de tecnologia da informação e dos resultados de suas práticas de combate à incitação ilegal ao ódio.

Ainda em 2016, saíram os primeiros resultados dessa avaliação, onde podemos observar dois dados interessantes. O primeiro foi sobre o principal motivo reportado às empresas como discurso de ódio: o discurso antissemita (que representou 27,3% dos casos). E depois, a quantidade média de conteúdo removido da Internet, onde 28% dos pedidos foram atendidos, sendo o *Youtube* aquele que atendeu a 48,5% - praticamente a metade - dos pedidos de remoção de conteúdo tido como ofensivo pelos usuários e observadores (COMISSÃO EUROPEIA, 2016).

Sem embargo, na quarta revisão publicada em fevereiro de 2019, o resultado foi um alerta: 72% do conteúdo alegadamente de incitação ao ódio foi eliminado das plataformas. No entanto, este resultado foi considerado extremamente positivo pela Comissão, não obstante ter reconhecido que as empresas de Tecnologia da Informação ainda “precisam ser mais transparentes nas notificações e nas remoções de conteúdo” (COMISSÃO EUROPEIA, 2019).

Finalmente, na quinta avaliação de resultados, a Comissão concluiu que o número permaneceu estável já que em 71% dos casos denunciados como *hate speech* (ou discurso de ódio), o conteúdo foi, de fato, removido das plataformas (COMISSÃO EUROPEIA, 2020).

Diante do exposto, analisados os fatos relatados nos referidos documentos europeus, entendemos o oposto daquilo que conclui a Comissão. É dizer, a falta de transparência na remoção de conteúdos publicados na Internet, como o próprio órgão relatou, demonstra apenas o insucesso de uma iniciativa como um “código de conduta para combater o discurso de ódio ilegal *on-line*”, além de que beira perigosamente à censura. Ademais, não se pode conceber que uma iniciativa que visa, com pouca ou nenhuma transparência, a remoção de opiniões da Internet, seja justificada, sem mais, como uma proteção à liberdade de expressão, sobretudo quando essa prática se apoia muito mais em políticas comerciais das próprias empresas privadas, do que nas normas constitucionais de qualquer Estado democrático.

Iniciativas como essa em análise, em última instância, representam a transferência da responsabilidade pela promoção de políticas de segurança, educação e tolerância nas redes sociais, para empresas privadas que não têm como norte a proteção e promoção de direitos fundamentais. Ademais, estas empresas demonstram o “sucesso” de seus números, porém eles apontam para a privação da liberdade alheia, sem grandes explicações e justificativas, o que se torna ainda pior quando observamos que os revisores de conteúdo, aqueles responsáveis pela exclusão da informação, são pessoas (ou robôs) sem nenhum conhecimento crítico sobre temas como liberdade de expressão, ponderação de direitos fundamentais, eficácia horizontal dos direitos fundamentais, dentre outros temas essenciais para a análise e conclusão pela eventual restrição de direitos dessa estirpe.

Em seu documento de fevereiro de 2019, com o objetivo de explicar como o Código de Conduta ajuda a combater o discurso de ódio ilegal na Internet, a Comissão Europeia comemorou que, em média, as empresas de Tecnologia estavam a avaliar 89% do conteúdo sinalizado em 24 horas, ante 81% no ano anterior. A taxa total de remoção permaneceu estável em cerca de 70%, o que, para a entidade, seria satisfatório, “pois o discurso de ódio não é fácil de definir” (COMISSÃO EUROPEIA, 2019). Ora, como ser satisfatória a remoção de 70% de

um conteúdo que sequer são capazes de definir? Estaria mesmo apenas o discurso de ódio sendo removido, ou neste número está contido algo mais? Como vemos, é difícil responder a essas questões, sobretudo quando não há transparência nas atitudes e nos dados das empresas e das autoridades públicas.

Resultado diferente foi aquele conquistado através da recente aprovação do Regulamento dos Serviços digitais proposto pela Comissão Europeia, em julho de 2022, o qual nos parece uma inovação legislativa mais promissora, coerente e alinhada aos desafios da sociedade atual.

Este referido documento institui um quadro claro e eficaz em matéria de transparência e responsabilidade das plataformas. Algumas das novas obrigações propostas pela legislação são: a presença dos direitos fundamentais como norteadores dos serviços prestados; a obrigação de prestar informações aos usuários; a instituição de mecanismos de reclamações e recursos contra as decisões das plataformas; a criação de um mecanismo de resolução extrajudicial dos litígios; maior transparência nos serviços prestados e; a criação de um código de conduta (COMISSÃO EUROPEIA, 2022).

Como vemos, a normativa europeia, sobretudo diante da inovação trazida pela Comissão e aprovada recentemente pelo Parlamento, parece estar próximo a trilhar um caminho mais seguro e justo para o problema da regulação do discurso nas redes sociais, uma vez que, além de ter o respeito aos direitos fundamentais como norte, tem exigido transparência na atividade comercial, bem como, a criação de mecanismos de solução de controvérsia com recurso ao contraditório. É, sem dúvidas, um modelo que pode ser seguido em outras realidades regionais.

## **O problema da remoção de conteúdo e as políticas comerciais**

A remoção de conteúdo *on-line* sob a justificativa de promoção de um ambiente saudável e pacífico de discussão, não pode ser usada como pano de fundo para a violação de direitos humanos. Os Estados não podem usar as



grandes empresas de comunicação da Internet como ferramentas para limitar expressões as quais eles próprios seriam impedidos de o fazer. É dizer, se os próprios Estados pretendessem monitorar, controlar e editar os assuntos discutidos no ambiente *on-line* da maneira como é feito pelas empresas privadas, certamente, tais atitudes seriam tidas, numa realidade democrática, como inconstitucionais.

Assim, ao estimular que as empresas privadas construam o seu modelo de negócios e adotem as suas políticas comerciais sem observância aos direitos fundamentais, esses Estados escusam-se de atitudes inconstitucionais ou antidemocráticas ao passo em que, na verdade, cobram que atitudes desta natureza sejam levadas a cabo por entes privados em seu lugar.

Toda regulação que esses Estados pretenderem exigir das empresas deve ser justificada de acordo com os preceitos norteadores do direito internacional (ONU, 2019). Neste sentido, como resta claro, os Estados não podem criar um dever de controle da “paz social” no ambiente *on-line* a ser levado adiante pelas empresas privadas, quando nem mesmo os próprios Estados são capazes de o fazer, sobretudo se esse controle advém de filtros de conteúdo (naquele sentido trazido por Dafne Keller (2019), que signifiquem censura prévia.

Com efeito, o fato é que essas grandes empresas que funcionam na Internet são detentoras de grande poder de controle social através, sobretudo, da linguagem. Como podemos conhecer pela imprensa e por notícias divulgadas nas próprias redes sociais, os executivos dessas empresas de comunicação se reúnem periodicamente para debater os problemas mais recentes em matéria de discursos de ódio, desinformação e outros conteúdos perturbadores (THE ECONOMIST, 2018). Eles dividem preocupações e buscam soluções que nem sempre passam pela observância aos direitos fundamentais. O fato é que as decisões tomadas nesses fóruns sobre a remoção de conteúdo, acabam servindo de instruções para milhares de revisores, aparentemente, sem nenhum compromisso com parâmetros mínimos de justiça.

Neste cenário, observamos que as empresas moderam cada vez mais o conteúdo publicado em suas plataformas e essa não é mais vista como uma atividade periférica, mas sim, como essencial para seus negócios, sendo inclusive, um pleito da própria opinião pública. Cada uma dessas empresas elabora diferentes políticas, mas com um conteúdo em comum: tornar suas plataformas menos “tóxicas” ao filtrar toda matéria publicada.

É notório que, hoje, um número pequeno de empresas consegue controlar o que milhões de pessoas ao redor do globo podem dizer, ouvir e ver. Para ilustrarmos essa discussão, é emblemático o caso do aplicativo de mensagens chinês *WeChat*, que conta com milhões de usuários e adota em suas políticas a proibição expressa do discurso de ódio. O problema é que o que a empresa entende por discurso de ódio é todo aquele conteúdo que, de fato, ou em sua “opinião razoável é odioso, hostil, abusivo, racial ou etnicamente ofensivo, difamatório, humilhante para outras pessoas, ameaçador, profano ou censurável” (WECHAT, 2022). Ou seja, todo e qualquer pensamento, segundo a “opinião razoável” da empresa, pode ser excluído, pois é ela quem decide quais conteúdos podem ser compartilhados, o que, inegavelmente, influencia a opinião pública à sua maneira.

## **O papel regulador das mídias sociais como entes não estatais e o problema da falta de transparência**

Diante do exposto, notamos que a maneira como as empresas tomam suas decisões, de forma pouco ou nada transparentes, a respeito de quais conteúdos serão retirados do ar, é alvo de muitas críticas. Por outro lado, elas têm muitas justificativas para atenuar estas críticas.

O *Facebook*, por exemplo, alega que a remoção de páginas é feita apenas após relatos de outros usuários. No conhecido caso Alex Jones, que foi banido de uma série de redes sociais, segundo essa empresa, a exclusão de sua conta se deu porque o autor estava fazendo apologia a violência e usando de linguagem desumana para descrever pessoas transexuais, muçulmanas e imigrantes. Já no



*YouTube*, no mesmo caso, a posição oficial foi a de encerrar a conta do usuário porque violaram as suas políticas contra o discurso de ódio e assédio e porque tentou escapar do cumprimento de suas regras (THE ECONOMIST, 2018). Em caso mais polêmico e mais recente, todas as principais plataformas resolveram banir definitivamente o então presidente dos Estados Unidos, Donald Trump, por incitar a violência provocada na invasão do Capitólio em 6 de janeiro de 2021 (BLOOMBERG; EURONEWS, 2021).

Com efeito, entendemos que quando a indústria da comunicação se limitava a pequenas *startups* jovens e vulneráveis, a exemplo do *Facebook*, que nasceu dentro de uma comunidade universitária com apenas alguns alunos usuários, onde todos se conheciam e onde se objetivava a manutenção das relações saudáveis dentro dessa pequena realidade; essa abordagem de monitorar o conteúdo postado ali era razoável. Contudo, hoje esta mesma plataforma conta com milhões de usuários em todo o mundo, o que demonstra que esses sítios se tornaram espaços centrais para a interação social e para todos os tipos de expressão de ideias.

É diante dessa realidade que a discussão sobre o papel regulador das mídias sociais como entes não estatais se coloca, já que, ao invés de promover o livre mercado de ideias na Internet, como se propunham inicialmente, as grandes empresas têm optado por mais vigilância e monitoramento dos conteúdos postados, ora por pressão da sociedade, ora por pressão regulatória dos Estados e ora por interesses econômicos próprios.

O *Facebook* e o *YouTube*, por exemplo, aumentaram drasticamente os recursos humanos e tecnológicos dedicados ao policiamento de suas plataformas nos últimos anos. O *Facebook* dobrou o número de funcionários e prestadores de serviços dedicados à "proteção e segurança", sendo grande parte deles "revisores de conteúdo". Já no *YouTube*, cerca de 10 mil pessoas trabalham, de forma direta ou indireta, na moderação do que é postado na plataforma (THE ECONOMIST, 2018). Como sabemos, essas empresas retiram milhões de postagens todos os

meses de cada plataforma, guiados por manuais de instruções espessos e vagos, que demonstram pouca clareza e segurança.

Com efeito, tanto investimento em segurança das informações além de se demonstrar inútil, já que notícias falsas sempre encontrarão uma maneira de serem divulgadas, eleva-se muito os custos das empresas que precisam investir cada vez mais em mão de obra e em tecnologia. Mark Zuckerberg, executivo-chefe do *Facebook*, já afirmou que, a longo prazo, o problema da moderação de conteúdo terá que ser resolvido com inteligência artificial, sob pena de ter de manter mais moderadores de conteúdo do que engenheiros em sua folha de pagamento (THE ECONOMIST, 2018).

O sistema de moderação do *YouTube* é semelhante, conta também com diretrizes publicadas para aquilo que é considerado por eles como aceitável, e ainda traz instruções detalhadas para seus funcionários revisores de conteúdo, que decidam rapidamente o que fazer com aquilo que foi sinalizado como falso ou ofensivo, e a maioria deles é feita por meio automatizado. O *Twitter* também usa inteligência artificial para identificar contas falsas e algum conteúdo inadequado, mas depende mais de relatórios de usuários que apontam, por exemplo, terem sofrido algum tipo de assédio ou intimidação.

Todas essas informações nos demonstram que se gasta muito dinheiro e energia em novas tecnologias e em capacitação humana para evitar a postagem de conteúdos ofensivos ou falsos, e muito pouco naquilo que resolve o problema da desinformação e da intolerância de forma sustentável: a educação midiática.

Neste sentido, muito se tem discutido sobre a alfabetização midiática, onde autores como Galeano apontam para a necessidade de alerta sobre os baixos percentuais, principalmente nos países latino-americanos, em termos de domínio de habilidades básicas e avançadas relacionadas com as tecnologias da informação. Tal fato pode ser devido às várias lacunas enfrentadas pela população desta região, tanto no acesso quanto no uso das tecnologias.

Conforme Galeano (2023), mostra-se necessário promover ações contra essas lacunas, que afetam as habilidades dos cidadãos para diferenciar

informações verdadeiras das falsas, para entender as implicações dos conteúdos que circulam tanto nos meios tradicionais de mídia, em sites e redes sociais por meio da Internet e/ou produzir suas próprias mensagens nesses ambientes.

Diante do exposto, fica evidente a problemática aqui discutida, bem como, a dificuldade de solução. Neste sentido, passamos a analisar outro ponto igualmente importante, qual seja, o problema do conceito de discurso de ódio.

## **A falta de clareza no conceito de discurso de ódio**

Como vemos, o trabalho de monitorar notícias pode mitigar os efeitos negativos da publicação de conteúdo inverídico, contudo, não impede a sua viralização e replicação (ALVES, 2021). Tal conduta tampouco impede que o problema se repita em outro local e que atinja outras vítimas, neste sentido, é preciso o empenho de gerações tolerantes e capazes de analisar criticamente uma opinião, de buscar outras fontes de informação e de identificar se aquilo se trata de um conteúdo mentiroso.

Sobre o problema do discurso de ódio, é notório que ele não pode ser admitido como o livre exercício do direito à manifestação do pensamento ou opinião, e, portanto, extirpado do discurso público, contudo, para tal, é essencial defini-lo de maneira a que todos o possam conhecer. A definição desse tipo de discurso deve, portanto, figurar publicamente nas políticas, regras e normas de todas as plataformas que pretendam oferecer transparência e segurança aos seus usuários de forma clara e que não restem dúvidas, e para isso, as empresas precisam defini-lo objetivamente.

Por outro lado, a presente discussão também deve passar pelo uso dos algoritmos nas redes sociais, se eles são capazes de aprofundar as divisões políticas e se facilitam a disseminação de desinformação e propaganda. É dizer, ao tentar maximizar o engajamento do usuário, os algoritmos maximizam a indignação dos demais? O próprio modelo de negócios adotado pelas grandes

empresas de comunicação, portanto, pode também ser susceptível, em última análise, de provocar o ódio e a divisão da sociedade (THE GUARDIAN, 2018).

Os critérios usados pelas plataformas, tanto em matéria de publicidade, como para moderar o conteúdo postado, não são nem objetivos e nem claros, e as próprias empresas, muitas vezes, assumem que o grande volume de conteúdo lhes faz cometer erros (THE NEW YORK TIMES, 2018). Geralmente a remoção de um texto ou imagem se dá em razão do número elevado de “denúncias” sobre o seu conteúdo, então, aquele que obtiver mais denúncias, será retirado mais rapidamente, o que nos parece perigoso, pois a adoção de um critério apenas quantitativo nessa política, pode representar, na verdade, uma imposição de opinião.

Isto exposto, percebemos que a falta de transparência em relação ao tratamento dos dados dos usuários e das próprias plataformas; à forma e às razões que levam à exclusão de certo conteúdo; e, principalmente, à falta de clareza sobre a definição de discurso de ódio; não por acaso, inflamam a discussão sobre a capacidade de manipulação da opinião pública por partes dessas grandes redes de comunicação na Internet.

Hoje, essas são algumas das instituições mais poderosas do mundo e moldam a opinião da sociedade e o discurso público de maneiras que nem os seus executivos conseguem explicar. É preciso, portanto, mais transparência e pesquisa, pois é apenas dessa forma que passaríamos a entender como as grandes plataformas da comunicação e a Internet estão afetando toda a sociedade.

Sem embargo, geralmente, os CEO’s destas empresas não confirmam qualquer ato de censura por parte de suas redes, e muito menos que monitorem o conteúdo postado seguindo qualquer viés ideológico (CNBC, 2018). Ainda assim, a pressão pela implementação de filtros de *upload* para discursos de ódio (e outros tipos de conteúdo), a nosso ver, não é uma boa solução, pois direciona as plataformas para a regulação e remoção também de conteúdo legal, gerando prejuízos incalculáveis para todos. O uso desses filtros aumenta o poder das

empresas com muito pouca ou nenhuma supervisão ou oportunidade de reparação. É dizer, sem uma legislação clara sobre discurso de ódio, deixar a regulação da liberdade de expressão nas mãos de empresas privadas pode ter consequências irreversíveis.

Na opinião do Relator Especial da ONU para a liberdade de expressão, os Estados deveriam buscar leis e políticas que pressionem as empresas a proteger a liberdade de expressão e combater formas legalmente definidas de discurso de ódio por meio de uma combinação de recursos, a saber, requisitos de transparência que permitam a supervisão e o escrutínio público, aplicação da lei nacional por autoridades judiciais independentes, bem como, outros esforços sociais e educacionais; de acordo com as diretrizes propostas na resolução do Conselho de Ação de Rabat e na resolução do Conselho de Direitos Humanos 16/18 (ONU, 2019).

## **A necessidade de regulação de conteúdo**

Diferente do que se demonstrava quando do surgimento das plataformas digitais, onde a promessa e a expectativa giravam em torno da construção de um ambiente de liberdade absoluta, o que tem se demonstrado, atualmente, é a necessidade de regulação do ambiente digital. Neste sentido, mostrar-se necessário, também, a proibição da publicação de certo conteúdo tido como ofensivo ou mentiroso, a exemplo dos discursos de ódio.

Contudo, nesta atividade, é fundamental que seja priorizada a observância daquilo que está estabelecido no Direito Internacional, é dizer, para além da necessidade, da legitimidade, da legalidade e da proporcionalidade de tais atos restritivos, é necessário que haja o estrito respeito aos direitos fundamentais, nomeadamente, a liberdade de expressão, assim como acontece com a regulação do discurso na vida fora da Internet.

Todo o aqui exposto nos leva a dois pontos de conclusão: primeiramente, se a desinformação, o discurso de ódio e o conteúdo ofensivo são tão difundidos

*on-line*, temos que ter em mente que estes podem ser efeitos do próprio modelo de negócios adotado nos últimos anos pelas empresas, que têm, cada vez mais, aberto espaço para publicidade e propaganda de qualquer material. Tal ponto não pode ser considerado irrelevante, pois, para vender mais e promover mais anúncios, os algoritmos dessas plataformas favorecem o conteúdo que, muitas vezes, pode ser ofensivo para alguns. Sabemos que as empresas de comunicação, quando vendem espaços publicitários, não estão preocupadas, exatamente, em proteger os usuários de ofensas ou de notícias falsas, pois o seu objetivo, aqui, é fazer negócios como qualquer outra empresa.

Em segundo lugar, também entendemos que ofensas, discurso de ódio e *Fake News* são efeitos colaterais do avanço tecnológico das comunicações sentido pela sociedade atual, pois diante do grande aumento de usuários dessas plataformas de comunicação, diante de meios mais interativos de troca de conteúdo nelas, e ainda, diante do fato de que as pessoas estão protegidas atrás de um computador onde elas sequer precisam dizer quem são ou assumir responsabilidades, é natural sentirem-se mais à vontade em propagar notícias sem o cuidado de checar sua veracidade ou de não provocar ofensas. Sem embargo, obviamente, o fato de estes serem efeitos naturais da atualidade, não tornam os discursos de ódio algo lícito. Tais discursos podem ser censurados, desde que haja uma definição clara de seu sentido.

Assim, em paralelo à promoção da educação e da tolerância nas interações sociais de comunicação, das quais as grandes empresas da Internet deveriam ser grandes promotoras, elas devem também, adotar em suas políticas regras claras que informem ao usuário quando este esteja a abusar de seu direito à livre manifestação do pensamento.

## **A regulação privada do discurso público**

A Lei Alemã NetzDG, de junho de 2017 que, dentre outros temas, trata da questão da remoção de conteúdo, é a mais conhecida deste tipo, mas vários



outros ordenamentos jurídicos têm optado pela adoção de regras que exigem das empresas de Internet a exclusão de falas ou textos considerados “manifestamente ilegais” dentro de um período temporal específico, normalmente, 24 horas (GERMAN LAW ARCHIVE, 2017).

Blanco de Moraes (2018), ao dedicar-se ao estudo desta lei, afirma que o poder público alemão transformou as grandes firmas de comunicação em verdadeiras gestoras da rede, pois, sob pena de pesadas multas, elas são responsáveis pela remoção de conteúdos “ilícitos” como *fake news* e *hate speech* – este último, como alerta o autor, ainda sofre o agravante de ter sido entendido num sentido perigosamente lato, portanto, ao fim e ao cabo, quem vai determinar se estaremos mesmo diante de um discurso odioso ou não, será uma empresa privada com fulcro no ganho pessoal, não o Estado que se pauta em todas as relações pelo respeito aos direitos fundamentais.

Assim, diante do novo cenário em que a comunicação social se encontra e diante da impossibilidade de controlar tudo o que é dito na Internet, o Estado passa essa responsabilidade para as empresas privadas, e então, agora são elas quem decidem o que é um discurso de ódio, o que fere a dignidade humana, o que é notícia falsa, e o que, portanto, deve ser banido do ambiente *on-line*. Se fosse o Estado a tomar atitudes dessa estirpe, ainda que sob a justificativa de proteção de minorias contra ofensas propagadas neste ambiente, seriam tidas, no mínimo, como medidas inconstitucionais e antidemocráticas.

Portanto, a finalidade das empresas não é fazer um exercício de ponderação entre os direitos fundamentais ali envolvidos. Muito menos elas têm como norte a observância aos parâmetros da legalidade, legitimidade, necessidade e proporcionalidade, como conferido pelo Direito Internacional. Ainda pelo contrário, como as empresas são pautadas pelo lucro, diante de uma exigência de remoção de conteúdo num prazo de 24 horas, sempre que houver dúvida sobre a interpretação de um discurso ou veracidade de uma notícia, a fim de evitar uma punição financeira (que, no caso alemão, pode chegar a 5 milhões de euros) a empresa certamente optará pela retirada do conteúdo, sobretudo

porque se ela cometer algum equívoco nessa atitude, nenhum tipo de responsabilização recairá sobre si. Então, em caso de dúvida, é muito mais vantajoso para a plataforma a opção pela censura. A prioridade das empresas, como resta óbvio, é cumprir a determinação da legislação aplicável e retirar o conteúdo ainda dentro do prazo estipulado, evitando punições, inclusive financeiras.

Neste sentido, o Relator Especial da ONU para a liberdade de expressão afirmou que a lei alemã aqui estudada (*Netzdurchsetzungsgesetz*) é “problematicamente vaga”, e embora deva ser entendida como um esforço de boa-fé para lidar com a preocupação generalizada com o ódio *on-line* e suas consequências *off-line*, a falha na definição de termos essenciais, a exemplo do discurso de ódio, dá margem à justas críticas cujas opiniões apontam para a incompatibilidade da lei com os *standards* do Direito Internacional dos Direitos Humanos (ONU, 2019). É dizer, a tentativa de inovação legislativa, aqui, vem antes de ser solucionada uma questão preliminar: a de definir o que é o discurso de ódio.

Como bem demonstra John Stuart Mill, há um sentimento na mente de cada pessoa de que todos deviam ser obrigados a agir da mesma forma que ela e aqueles com quem simpatiza, pois, as opiniões das pessoas sobre o que é louvável ou condenável, são afetadas por todas as diversas causas que influenciam os seus desejos ou sentimentos em relação à conduta dos outros, causas que são tão numerosas quanto as que determinam os seus desejos sobre qualquer outro assunto, como sua razão, seus preconceitos, superstições, seus sentimentos e, mais comumente, os seus interesses próprios, sejam eles legítimos ou não (MILL, 2015). Com efeito, tal afirmativa reforça a necessidade de que as condutas e conceitos sejam claros, para apenas assim, pensarmos em limitar a liberdade de expressão.

Portanto, a questão que se coloca não é sobre a hipótese de exclusão do conteúdo postado, pois a liberdade de expressão não é absoluta. O que tem sido motivo de preocupação é o fato de que estas condutas censoras estão a ser

levadas a cabo sem que haja legislação que as justifique, nem tampouco, clareza e transparência em sua condução.

Como podemos observar no debate público, são precisamente essas empresas as responsáveis pela rápida disseminação de praticamente todas as notícias as quais temos acesso, muita delas em tempo real, e por isso, a influência causada por elas aos seus leitores e usuários é mais do que evidente, contudo, não deve caber a elas a última palavra sobre aquilo que poderá, ou não, ser lido, visto, assistido e ouvido pela sociedade.

Sobre este aspecto, Stuart Mill diz que a censura privada é tão má quanto a pública, pois há um limite à interferência legítima da opinião coletiva na independência individual e encontrar esse limite - e protegê-lo contra transgressões - é tão indispensável para o bom estado das relações humanas, como a proteção contra o despotismo político (MILL, 2015).

## **Conclusão**

O que tudo o aqui exposto significa para os contornos da liberdade de expressão na era das mídias sociais não é imediatamente claro, pelo contrário, é uma área cinzenta entre a necessidade de criação de uma estrutura legal na qual as empresas de comunicação possam ser responsabilizadas por suas decisões editoriais, por um lado; e por outro, que não prejudique a livre iniciativa e sua autonomia para filtrar informações perigosas e comprovadamente falsas e/ou enganosas.

Porém, se pretendem realmente resolver o desafio da moderação de conteúdo, as grandes plataformas devem começar por impor mais rigidez em seus próprios modelos de negócios, pois ao que parece, estas empresas agem rapidamente em monitorar o conteúdo supostamente ofensivo e falso quando denunciado por seus usuários, no entanto, esta rapidez e cuidado não são verificados da mesma maneira quando o suposto conteúdo abusivo decorre de

clientes que adquirem o seu espaço para divulgar ou promover seus próprios interesses.

Isso exposto, defendemos que, assim como no ambiente *off-line*, as plataformas digitais não só podem, como devem excluir as publicações que contenham discursos de ódio, contudo, é primordial que seja definido claramente o que representa este tipo de discurso e quem são suas vítimas, pois a falta de clareza e a vagueza com que o termo é tratado, não traz outra coisa senão insegurança jurídica e ameaça ao livre exercício de um direito elementar como a liberdade de expressão, o que é inaceitável numa democracia.

Não há aqui razão para defender outros limites à liberdade de expressão somente porque o discurso ocorre no ciberespaço. Essas novas ferramentas que trouxeram com elas uma nova forma de comunicação, de informação e de troca dos mais variados tipos de conteúdo, representam um caminho sem volta para a sociedade atual, e, portanto, cabe a nós apontarmos soluções para os problemas comuns que surgem de toda novidade. No entanto, a censura, sem critérios normativos rígidos e transparentes, certamente não é uma hipótese.

Neste sentido, sem prejuízo do debate sobre a regulação, deve-se exigir dessas companhias, primeiramente, transparência na elaboração de suas políticas, e depois, o compromisso de que, no trabalho de moderação de conteúdo, sejam assegurados os princípios que regem o Direito Internacional dos Direitos Humanos, pois, um ambiente virtual livre e seguro denota um espaço onde os direitos dos quais seus usuários dispõem são exatamente os mesmos daqueles dos quais gozam *off-line*.

## Referências

ABDO, A. **Facebook is shaping public discourse. We need to understand how.** *The Guardian*, 15 de setembro de 2018. Disponível em: <https://www.theguardian.com/commentisfree/2018/sep/15/facebook-twitter-social-media-public-discourse> . Acesso em: 26 de setembro de 2018.

ALVES, M. **Clones do YouTube: replataformização da irrealdade e infraestruturas de desinformação sobre a Covid-19.** *Revista Fronteiras*. Porto Alegre : Unisinos, 2021.

BLANCO DE MORAIS, C. **Redes sociais vs. Media profissional: “Vídeo, mentiras e hate speech”**. *Público*, Lisboa, 8 de Dezembro de 2018. Disponível em: <https://www.publico.pt/2018/12/08/sociedade/opiniao/redes-sociais-vs-media-profissional-video-mentiras-hate-speech-1853221#gs.OHjL1uEZ> . Acesso em: 8 de Fevereiro de 2019.

BROMWICH, Jonah Engel. **YouTube Cracks Down on Far-Right Videos as Conspiracy Theories Spread**. *The New York Times*, 3 de março de 2018. Disponível em: <https://www.nytimes.com/2018/03/03/technology/youtube-right-wing-channels.html> . Acesso em: 26 de setembro de 2018.

COMISSÃO EUROPEIA . **Code of conduct on countering illegal hate speech online**. 31 de maio de 2016, § 2º. Disponível em: [http://ec.europa.eu/justice/fundamental-rights/files/hate\\_speech\\_code\\_of\\_conduct\\_en.pdf](http://ec.europa.eu/justice/fundamental-rights/files/hate_speech_code_of_conduct_en.pdf) . Acesso em: 6 de fevereiro de 2019.

\_\_\_\_\_. Comunicado de imprensa. **Combate à incitação ilegal ao ódio em linha – Código de conduta da EU garante uma resposta rápida**. Bruxelas, 4 de fevereiro de 2019, §§ 2º e 5º. Disponível em: [http://europa.eu/rapid/press-release\\_IP-19-805\\_pt.htm](http://europa.eu/rapid/press-release_IP-19-805_pt.htm) . Acesso em: 6 de fevereiro de 2019.

\_\_\_\_\_. **Countering illegal hate speech online: 5<sup>th</sup> evaluation of the Code of Conduct**. Junho de 2020, p. 2. Disponível em: [https://ec.europa.eu/info/sites/info/files/codeofconduct\\_2020\\_factsheet\\_12.pdf](https://ec.europa.eu/info/sites/info/files/codeofconduct_2020_factsheet_12.pdf) . Acesso em: 23 de setembro de 2020.

\_\_\_\_\_. **Regulamento Serviços Digitais**. Disponível em: [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment\\_pt](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_pt) . Acesso em: 16 de novembro de 2022.

CONSELHO. **Decisão-Quadro 2008/913/JAI, de 28 de novembro de 2008, relativa à luta contra certas formas e manifestações de racismo e xenofobia**. Jornal Oficial da União Europeia, L328/55, de 6.12.2008. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32008F0913&from=EN> . Acedido em: 6 de fevereiro de 2019.

DONALD TRUMP'S Twitter ban is “problematic”, says Angela Merkel. **Euronews**. 12 de janeiro de 2021. Disponível em:

<https://www.euronews.com/2021/01/12/donald-trump-s-twitter-ban-is-problematic-says-angela-merkel> . Acedido em: 14 de janeiro de 2021.

FRIER, S. **Bans on Parler and Trump Show Big Tech's Power Over Web Conversation.** *Bloomberg*, 10 de janeiro de 2021. Disponível em: <https://www.bloomberg.com/news/articles/2021-01-11/parler-trump-bans-show-big-tech-s-power-over-web-conversation> . Acesso em: 14 de janeiro de 2021.

GALEANO, B. **Alfabetización Mediática y aprendizaje informal en Latinoamérica.** Revista do Programa de Pós Graduação em Comunicação. Juiz de Fora : UFJF, 2023.

GERMAN LAW ARCHIVE. **Network Enforcement Act (Netzdurchsetzungsgesetz, NetzDG).** Setembro, 2017, art. 1º, seção 4 (2). Disponível em: <https://germanlawarchive.iuscomp.org/?p=1245> . Acesso em 7 de outubro de 2020.

HIGGINS, T. **Supreme Court agrees to hear a case that could determine whether Facebook, Twitter and other social media companies can censor their users.** *CNBC*, 16 de Outubro de 2018. Disponível em: [https://www.cnbc.com/2018/10/16/supreme-court-case-could-decide-fb-twitter-power-to-regulate-speech.html?\\_source=sharebar%7Cemail&par=sharebar](https://www.cnbc.com/2018/10/16/supreme-court-case-could-decide-fb-twitter-power-to-regulate-speech.html?_source=sharebar%7Cemail&par=sharebar) . Acesso em: 1 de Novembro de 2018.

KELLER, D. **Dolphins in the Net: Internet Content filters and the Advocate General's Glawischnig-Piesczek v. Facebook Ireland Opinion.** Stanford Center for Internet and Society : Stanford, 2019.

MILL, J. S. **Sobre a Liberdade.** Lisboa : Edições 70, 2015.

PARK, Melo. **How social-media platforms dispense justice.** *The Economist*, 6 de setembro de 2018. Disponível em: <https://www.economist.com/business/2018/09/08/how-social-media-platforms-dispense-justice?cid1=cust/ednew/n/bl/n/2018/09/6n/owned/n/n/nwl/n/n/LA/149526/n> . Acesso em: 18 de setembro de 2018.

UN. United Nations. Human Rights Council. **Report of the Special Rapporteur on the promotion and protection of the freedom of opinion and expression.** A/74/486. Genebra, 2019. Disponível em: [https://www.ohchr.org/Documents/Issues/Opinion/A\\_74\\_486.pdf](https://www.ohchr.org/Documents/Issues/Opinion/A_74_486.pdf) . Acesso em 6 de dezembro de 2019.



\_\_\_\_\_. **Report of the United Nations High commissioner for Human Rights on the expert workshops on the prohibition of incitement to national, racial or religious hatred.** A/HRC/22/17/Add.4. Genebra, 2013. Disponível em: <https://undocs.org/en/A/HRC/22/17/Add.4> . Acesso em: 19 de dezembro de 2019.

WECHAT. **Acceptable use policy.** Disponível em: [www.wechat.com/en/acceptable\\_use\\_policy.html](http://www.wechat.com/en/acceptable_use_policy.html) . Acesso em: 16 de novembro de 2022.

