

FACULDADE DAMAS DA INSTRUÇÃO CRISTÃ
CURSO DE DIREITO

CARLOS ANTÔNIO MENEZES DE ALBUQUERQUE

**REGULAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS: a moderação de conteúdo
e seus reflexos na liberdade de expressão em tempos de desinformação**

RECIFE

2022

CARLOS ANTÔNIO MENEZES DE ALBUQUERQUE

**REGULAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS: a moderação de conteúdo
e seus reflexos na liberdade de expressão em tempos de desinformação**

Monografia apresentada à Faculdade Damas da Instrução
Cristã como requisito parcial para obtenção do título de
Bacharel em Direito.

Orientador: Prof. Dr. Henrique Weil Afonso

RECIFE

2022

Catálogo na fonte
Bibliotecário Ricardo Luiz Lopes CRB-4/2116

A345r Albuquerque, Carlos Antônio Menezes de.
Regulação das plataformas de redes sociais: a moderação do conteúdo e seus reflexos na liberdade de expressão em tempos de desinformação / Carlos Antônio Menezes de Albuquerque. - Recife, 2022. 52 f.

Orientador: Prof. Dr. Henrique Weil Afonso.
Trabalho de Conclusão de Curso (Monografia - Direito) – Faculdade Damas da Instrução Cristã, 2022.
Inclui bibliografia.

1. Desinformação. 2. Liberdade de expressão. 3. Moderação de conteúdo. I. Weil, Henrique Afonso. II. Faculdade Damas da Instrução Cristã. III. Título

340 CDU (22. ed.)

FADIC (2022.1-004)

FACULDADE DAMAS DA INSTRUÇÃO CRISTÃ
CURSO DE DIREITO

CARLOS ANTÔNIO MENEZES DE ALBUQUERQUE

REGULAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS: a moderação de conteúdo e seus reflexos na liberdade de expressão em tempos de desinformação

Defesa Pública em Recife, __ de junho de 2022

BANCA EXAMINADORA:

Presidente

Examinador(a)

AGRADECIMENTOS

Agradeço à minha família pelo incentivo e paciência de sempre com mais essa aventura em busca de conhecimento, aos grandes amigos que ganhei e muito somaram nessa jornada (Carol Beltrão, Carlos Guerra, Ileana Nobrega, Léo Junqueira e Mariza Alpes), aos professores dedicados nessa grandiosa arte. Foram anos velozes e de maravilhosas surpresas. Conhecimentos, organização, ideias, sonhos, incansáveis noites, habilidades e amizades.

RESUMO

O presente trabalho faz uma análise da evolução da esfera pública, alterada pelas novas tecnologias de comunicação, como as plataformas de redes sociais, em que a disseminação da informação é instantânea e desconhece limitações geográficas. É nesse cenário de hiperinformação, que surgem problemas como a desinformação em massa, assim os Estados começam a discutir legislações para que o conteúdo trafegado não se desvirtue da legalidade. Dessa forma, este estudo confronta o poder alcançado pelas plataformas de redes sociais na condução da informação com as necessidades jurídicas para proteção à liberdade de expressão. Para isso, foi utilizada a metodologia descritiva, valendo-se de observações qualitativas, pela verificação da legislação, doutrina, jurisprudência, projetos de lei e documentação técnica, para se alcançar de maneira dedutiva a problemática estabelecida para regulação das plataformas de redes sociais e a preservação da liberdade de expressão afetada pela moderação de conteúdo. Neste sentido, a pesquisa analisou a experiência alemã para regulação das plataformas de redes sociais com a lei NetzDG e a iniciativa brasileira com o projeto de lei 2630/2020, e sobretudo o instituto da autorregulação regulada presente nestes dois dispositivos. Verificou-se que a legislação alemã já se encontra em execução trazendo resultados satisfatórios com a implantação da autorregulação regulada e não há relatos que impliquem em afetações à liberdade de expressão, podendo servir de orientação a outros Estados. Já o projeto de lei brasileiro, encontra-se com dificuldades para sua aprovação, onde uma série de ementas podem descaracterizar a ideia original, precisando assim, de maior debate para que exista uma convergência política sobre o tema.

Palavras-chave: Desinformação. Liberdade de Expressão. Moderação de Conteúdo.

ABSTRACT

The present work analyzes the evolution of the public sphere, changed by new communication technologies, such as social networking platforms, in which the dissemination of information is instantaneous and ignores geographical limitations. It is in this hyper-information scenario that problems such as mass disinformation arise, so States begin to discuss legislation in order the traffic content does not deviate from legality. In this way, this study confronts the power achieved by social media platforms in conducting information with the legal needs to protect freedom of expression. For this, a descriptive methodology was used, using qualitative observations, by verifying legislation, doctrine, jurisprudence, bills and technical documentation, to reach in a deductive way the problems established for the regulation of social network platforms and the preservation of freedom of expression affected by content moderation. In this sense, the research analyzed the German experience for the regulation of social media platforms with the NetzDG law and the Brazilian initiative with the bill 2630/2020, and especially the institute of regulated self-regulation present in these two devices. It was verified that the German legislation is already in execution, bringing satisfactory results with the implementation of regulated self-regulation and there are no reports that imply in affectations to freedom of expression, being able to serve as a guide to other States. The Brazilian bill, on the other hand, is having difficulties for its approval, where a series of additional changes can de-characterize the original idea, thus needing greater debate so that there is a political convergence on the subject.

Keywords: Disinformation. Freedom of Speech. Content Moderation.

SUMÁRIO

1	INTRODUÇÃO	7
2	A ESFERA PÚBLICA E AS PLATAFORMAS DE REDES SOCIAIS	10
2.1	CONCEITOS IMPORTANTES SOBRE ESFERA PÚBLICA	10
2.2	MUDANÇAS NOS MEIOS DE COMUNICAÇÃO EM MASSA NA INTERNET	12
2.3	A INTERNET E AS PLATAFORMAS DE REDES SOCIAIS	16
3	LIBERDADE DE EXPRESSÃO, DESINFORMAÇÃO E MODERAÇÃO DE CONTEÚDO EM CONTEXTO.....	20
3.1.	A LIBERDADE DE EXPRESSÃO NO CENÁRIO BRASILEIRO	21
3.2.	O PROBLEMA DA DESINFORMAÇÃO EM MASSA	25
3.3	MODERAÇÃO DE CONTEÚDO EM PLATAFORMAS DE REDES SOCIAIS	27
4	REFLEXÕES SOBRE A REGULAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS.....	32
4.1.	A EXPERIÊNCIA ALEMÃ E A AUTORREGULAÇÃO REGULADA	33
4.1.1	A autorregulação regulada.....	37
4.2.	A INICIATIVA DO PL2630/2020.....	40
5	CONCLUSÃO.....	47
	REFERÊNCIAS.....	49

1 INTRODUÇÃO

A Internet é, sem sobra de dúvidas, uma das maiores invenções da humanidade e a partir dela, inúmeros avanços foram sendo concretizados, permitindo que a própria humanidade evoluísse em aspectos econômicos, sociais e culturais. Houve o desenvolvimento de um mundo virtual que, nos dias de hoje, se mistura ao mundo real. Faz parte do cotidiano o envio de uma simples mensagem de bom dia para um grupo da família por meio de aplicativo de mensageria como também é possível a realização de uma cirurgia a milhares de quilômetros de distância entre o médico e o paciente.

Dentre essas aplicações da Internet, surgiram as plataformas de redes sociais, como o *Facebook* e *Twitter*, que além de apenas juntar, virtualmente, pessoas com interesses em comum, são meios de comunicação em massa, em que cada usuário é um potencial produtor de conteúdo e sua informação pode chegar instantaneamente a milhões de pessoas em qualquer lugar do mundo. Pode-se falar que há uma mudança no trato da informação, uma transformação da Esfera Pública, inicialmente, pensada por Habermas, que parte, modernamente, de fontes de informação e formação de opinião por meio dos jornais impressos com redação para fontes descentralizadas produzidas por qualquer pessoa em um ambiente virtual.

Diversos são os temas que podem ser abordados nas redes: desde desinformação e discursos de ódio travestidos de liberdade de expressão, como também discursos legítimos que são censurados e removidos das plataformas. No âmbito tecnológico e do direito, muito se tem debatido sobre a transparência das plataformas na moderação de conteúdo e o quanto isso pode influenciar o debate político e nas Democracias.

Fatos recentes e amplamente divulgados, como a tentativa de interferência nas eleições Norte Americanas de 2016 associadas a mineração de dados dos usuários do Facebook, a invasão do congresso americano em 2021 e os atos antidemocráticos realizados no Brasil, mostram que as redes sociais são ferramentas importantes no cenário político e precisam receber as devidas atenções para que a sociedade possa caminhar de maneira sóbria para uma cooperação saudável.

Neste aspecto, é preciso que esse processo seja confrontado com uma visão do mundo do Direito, sobretudo, para que seja garantida a proteção à direitos fundamentais como a liberdade de expressão, direito à informação, à privacidade, mas, também, que se consolidem instrumentos de proteção da própria democracia.

Nesse contexto, já existem iniciativas legais para a moderação de conteúdo, como é o caso da Alemanha, pioneira no tratamento dessa nova sociedade das plataformas (*Netzdurchsetzungsgesetz, NetzDG*). A União Europeia já trabalha com a elaboração do *Digital Service Act* para proteção dos direitos fundamentais nas redes. No Brasil há o Marco Civil da Internet que é considerada uma legislação robusta, porém não trata especificamente a regulação das plataformas redes sociais. Novos projetos de lei com essa temática foram propostos, como é o caso do PL 2630/2020, conhecido como PL das *Fake News* que busca instituir a lei brasileira de liberdade, responsabilidade e transparência na internet e o PL3227/2021 que busca alterar o Marco Civil da Internet promovendo uma lista de itens passíveis moderação.

Diante deste cenário de várias visões de mundo amplificadas por câmeras virtuais de eco e que podem se valer de técnicas de desinformação para sobrepor o legítimo exercício da liberdade de expressão, bem como desestruturar as instituições democráticas, de que forma a regulação das plataformas de redes sociais, por meio da moderação de conteúdo, pode ser realizada para uma maior proteção aos direitos fundamentais, em especial, à liberdade de expressão?

Para condução de uma regulação no cenário de profunda mudança na esfera pública é preciso que exista uma sinergia entre o Privado, com uma autorregulação em que sejam estabelecidas regras claras para utilização da plataforma de rede social e remoção de conteúdo, e o Estado, envolto em seu interesse público estabelecendo limites bem definidos. Só assim, pode-se alcançar uma proteção de direitos fundamentais de maneira mais robusta.

Este trabalho tem como objetivo geral analisar a importância de se confrontar o poder alcançado pelas plataformas de redes sociais na condução da informação com as necessidades jurídicas para proteção à liberdade de expressão. Especificamente busca-se: a) analisar a transformação da esfera pública nos tempos atuais; b) abordar à liberdade de expressão no cenário brasileiro e a desinformação em massa via plataformas de redes sociais bem como as estratégias de moderação de conteúdo e c) refletir, por meio de uma análise comparativa de institutos legais, sobre a necessidade de uma regulação para a moderação de conteúdo e responsabilidades das plataformas, sobretudo a experiência alemã e a iniciativa do PL2630/2020.

A metodologia utilizada para condução deste estudo é a descritiva, valendo-se de observações qualitativas, pela verificação da legislação, doutrina, jurisprudência, projetos de lei e documentação técnica, para se alcançar de maneira dedutiva a problemática estabelecida para regulação das plataformas de redes sociais e a preservação da liberdade de expressão afetada pela moderação de conteúdo.

A distribuição temática deste trabalho é realizada com a seguinte forma: o primeiro capítulo trata de uma análise da transformação da esfera pública com o advento da Internet e mais especificamente no contexto das plataformas de redes sociais como uma espécie de púlpito virtual de alcance planetário para debates de interesse público.

O segundo capítulo aborda a liberdade de expressão sob a ótica brasileira, conceitua os principais elementos do problema da desinformação em massa estabelecendo, estabelecendo classificações e sua estruturação em campanhas nas plataformas de redes sociais, bem como analisa quais as abordagens utilizadas pelas empresas na moderação de conteúdos de terceiros disponibilizados em suas plataformas de redes sociais, assim como relaciona os principais problemas sobre os aspectos da liberdade de expressão atingida por estas moderações.

Por último há um capítulo que faz uma reflexão sobre a necessidade de regulação das plataformas de redes sociais buscando analisar a experiência alemã e o PL2630/2020 enfatizando a utilização da autorregulação regulada.

2 A ESFERA PÚBLICA E AS PLATAFORMAS DE REDES SOCIAIS

Para que se possa compreender o cenário atual e o fenômeno da influência das plataformas de redes sociais, é preciso que se observe um conceito abstrato que formula a organização de como a informação, ideias, cobranças e questionamentos entre sociedade e Estado se entrelaçam. Essa interação tem mudado ao longo dos anos e principalmente com aparecimento de novas tecnologias e meios de se transmitir a informação. Sob esse aspecto é que as plataformas de redes sociais, instituições privadas, modificam o patamar estruturante de como se consome e se produz a informação.

A virtualização do ambiente comunicacional traz inúmeros desafios, entre eles, o entendimento das novas formas de interação política ao qual estão sujeitas as sociedades contemporâneas. O poder de influência da sociedade, potencializado pelas redes sociais, pode alterar a maneira de como o ecossistema político atua e se comporta aos sinais emitidos pelo povo, porém, a superficialidade dos debates públicos permeadas pela instantaneidade das informações, podem levar a problemas no ambiente democrático.

Desta forma, este capítulo, tem como tema a discussão sobre conceitos de esfera pública e sua modificação estrutural, principalmente para que se possa estabelecer alguns aspectos importantes dos meios de comunicação em massa sob a infraestrutura da Internet, quando se observam a velocidade e alcance da informação trafegada.

Nesse contexto de mudança, é necessário, também, que sejam apresentados dados de algumas das plataformas de redes sociais, principalmente informações relevantes sobre a capilaridade e vulto econômico que alcançaram.

2.1 Conceitos importantes sobre Esfera Pública

Inicialmente, é preciso observar o conceito e evolução da esfera pública, que foi profundamente analisado na obra “Mudança Estrutural da Esfera Pública” de Jürgen Habermas e que serve de fonte para algumas passagens deste texto (HABERMAS, 2014). Nela o autor realiza uma investigação complexa passando por elementos históricos, sociais e institucionais de modo a trazer à tona informações valiosas para o entendimento de como se comporta o fluxo das informações e o debate público.

Estabelece delimitações de entendimento do que chama da esfera pública burguesa extraído de um contexto que é reflexo do desenvolvimento alemão, inglês e francês do século

XVIII e início do século XIX. A esfera pública burguesa trazia a manutenção interesses coletivos, mas não de maneira geral, pois a participação se dava de maneira restrita à determinadas classes. De toda sorte, ainda que voltada para uma camada social específica, era possível se reconhecer como lugar de argumentações e expressões sociais de interesse público (HABERMAS, 2014).

O autor observou inúmeros aspectos importantes na mudança estrutural da esfera pública sob o ponto de vista de sua transformação no Estado de bem-estar social e da sua modificação nas estruturas de comunicação, estas ocasionadas pelo surgimento dos meios de comunicação em massa. Este último tema se alinha mais aos objetivos deste trabalho, pois o aumento do fluxo das informações é um fator importante para se entender a dinâmica social estabelecida atualmente.

HABERMAS (1997) sustenta que a esfera pública pode ser entendida como uma estrutura intermediária que faz a conexão entre o sistema político e setores privados do mundo da vida. Analisa, também, que esta age para assegurar de maneira institucional o vínculo entre a lei e opinião pública. Extrai-se, assim, a esfera pública, como uma abstração de tráfego de informações, o *locus* onde há o fluxo da comunicação, permitindo que a opinião pública seja formulada e utilizada como organização do sistema político para direcionamento social.

Adiantada a linha do tempo, houve declínio do modo de vida da sociedade burguesa no final do século XIX dando-se início uma cultura de massas com os novos meios de comunicação, efeitos da modernidade industrial. Esse novo modelo de esfera pública da classe média com acesso aos meios de comunicação é caracterizado pelo pluralismo de grupos categorizados por temas e organizações da sociedade, como partidos políticos, emissoras de rádios, igrejas. Há uma espécie de esfera pública pluralista de grupos.

Posteriormente, surge uma esfera pública fragmentada em rede com grande desagregação do fluxo de comunicação. Alterações no modo de vida, extremamente dinâmico, em uma cultura formada em uma sociedade pós-industrial. As opiniões crescem e se alteram rapidamente (VESTING, 2020, p. 194-210).

A esfera pública vem sofrendo uma transformação estrutural, principalmente na maneira de integração entre sociedade e Estado. Os espaços de debates públicos passaram de lugares como bares, teatros e praças para jornais, periódicos e televisão, meios de comunicação em massa. Habermas descreve essa transformação da seguinte forma:

(...) A infraestrutura da esfera pública se modificou com as formas de organização, comercialização e consumo de uma produção de livros profissionalizada e ampliada, orientada por uma nova classe de leitores, e com uma imprensa de jornais e periódicos

cujo conteúdo também se modificou. A mudança estrutural da esfera pública ocorre mais uma vez com o surgimento dos meios de comunicação de massa eletrônicos, com a nova relevância da propaganda, com a crescente fusão de entretenimento e informação (...) (HABERMAS, 2014 p. 57)

Percebe-se que a transformação da infraestrutura da esfera pública possui uma relação como a evolução tecnológica. O debate é transportado de cafés e praças com um alcance limitado para meios em que a informação chega a mais e mais pessoas. Com a criação de jornais de grande circulação e a invenção do rádio e da televisão, a passagem do conhecimento foi ainda mais alargada. Com a Internet, a circulação da informação atingiu outro patamar, tanto em perímetro quanto em velocidade, pois pode chegar a qualquer lugar do planeta instantaneamente.

Para este trabalho, a esfera pública trata-se de uma dimensão onde a informação circula e o debate público é estabelecido para formulação de uma opinião pública que realiza a interação entre a sociedade e Estado.

2.2 Mudanças nos meios de comunicação em massa na Internet

Realizada a discussão sobre conceitos de esfera pública, é preciso que se exponham algumas características da infraestrutura que suportam o fluxo de informações. Pois, com a evolução tecnológica, foram surgindo os meios de comunicação em massa como o rádio, televisão e mais recentemente a Internet, que conduziram uma transformação no fluxo das informações consumidas em todo o planeta. Quanto maior o alcance da informação, maior é possibilidade de se disseminar ideias, e isso não foge ao campo político.

A debate público de logo alcance, com os meios de comunicação em massa, estabelece um campo de atuação político tanto para cobrança no sentido sociedade para o Estado, assim influenciando as decisões de cunho público, como também, há o caminho inverso, do Estado para sociedade na disseminação de sua intenção de organização social.

Essa interação existe desde a concepção das sociedades-Estado, o debate político para a organização social é algo comum, porém com a avanço tecnológico, proporcionando também a mudança estrutural da esfera pública, conduziu a discussão a outro patamar.

É interessante observar a direção do fluxo de informações nestes meios, enquanto os grandes jornais, rádios e televisão, a informação sai do emissor aos receptores de maneira unidirecional, na Internet, o emissor e receptor podem instantaneamente trocar de papel, e a

informação é trafegada de maneira multidirecional. Há a formação de uma espécie de malha comunicacional que se entrelaça por todo o espectro da informação.

Algumas características importantes dessa forma de comunicação podem ser observadas na formação da malha. São elas: instantaneidade, descentralização e fragmentação. A primeira, a informação, deixa de ser confrontada com a variável tempo, pois ao enviar uma mensagem pela Internet, a comunicação entre o emissor e receptor procede forma assemelhada a um diálogo no mundo físico, é em tempo real. A descentralização ocorre, pois na malha comunicacional, a direção da informação, pode ser a que se pretende realizar, um para um, um para muitos e muitos para muitos, o sentido é livre. A terceira característica, a fragmentação, permite que apenas parte de fatos circulem de maneira a se estabelecer narrativas as quais se pretende direcionar o discurso e moldar uma nova realidade.

Sob ponto de vista do Direito há um fator importante a ser ressaltado, que é o da responsabilidade da informação. Nos jornais impressos e televisão, há a figura do redator que faz as revisões necessárias nas matérias e textos produzidos, dessa forma, passando por um filtro de adequação. Observava-se uma centralidade na produção do conhecimento social, modelo que se apresenta como sociedade das organizações. Com a chegada e popularização da Internet, a geração da informação sai de uma vinculação legal de uma estrutura de redação para uma produção de informação generalizada sem vínculo revisional provocando uma descentralização informacional (ABBOUD; CAMPOS, 2020, p. 121-141).

Essa descentralização da produção da informação é uma das características da sociedade atual. Cada indivíduo, usando as plataformas de redes sociais, é um potencial produtor e consumidor da informação. Essa informação pode ser instantaneamente distribuída para milhões de pessoas em qualquer parte do mundo. A esfera pública como meio de interlocução entre sociedade e Estado se pluraliza ainda mais, torna-se mais dinâmica a forma de se atuar na disseminação de ideias e narrativas vinculadas a interesses próprios de indivíduos ou de grupos.

Essa facilidade promovida pela Internet e uso das redes sociais, proporcionou inúmeros benefícios para que mais e mais pessoas pudessem participar de maneira mais ativa das discussões sociais, políticas e econômicas. Dos mais diversos temas, dispostos nas redes, cabe observar que mais recentemente, o campo político ganhou notoriedade nas mídias pelo uso das redes sociais para se ter uma comunicação direta entre o político e a sociedade (VESTING, 2020, p. 194-210).

As campanhas políticas antes concentradas nas ruas, rádios e televisões, agora são realizadas, também, de maneira virtual nas mais diversas plataformas de redes sociais. A

velocidade da informação é um fator importante para que as plataformas políticas de um candidato sejam conhecidas rapidamente por uma quantidade expressiva de eleitores. Há também a possibilidade de se apresentar aos eleitores, em uma comunicação direta, o que cada político eleito tem realizado.

Esse mecanismo de interação política e sociedade tem sido bastante utilizado nos últimos anos, mais e mais atores políticos tem se valido das plataformas digitais para manter o seu capital político atento, mapeando cada ação realizada.

Uma análise feita por Campos (2017) sobre como transformação da esfera pública motivou uma nova lei alemã para Internet fez algumas considerações sobre este momento de uso das redes sociais na atuação política em especial após as eleições americanas de 2016 e o *Brexit*¹.

(...) Se anteriormente na sociedade das organizações a informação sobre assuntos públicos era gerada em grandes redações, com muitos profissionais especializados, na sociedade de redes a informação "perde a origem". De fato, a internet abriu a possibilidade de profissionais, que não se enquadravam nas organizações, de "gerenciar" seu próprio público sem a necessidade de associação formal à uma organização. Isso foi um ganho para a pluralidade informacional da esfera pública. Mas por outro lado, patologias experimentadas nas últimas eleições pelo globo em que se constatou que a maior parte da informação acessada era informação falsa preocuparam os que tomam o processo democrático como um processo correlato ao acesso à informação (em detrimento ao acesso à desinformação). (...)

Essa nova esfera pública, com uma infraestrutura tão dinâmica como as redes sociais, potencializa o fluxo de informações desde a sua produção até o seu consumo. Há um ganho expressivo na quantidade de informação trafegada, porém é possível questionar-se a qualidade da informação que chega aos receptores, pois diferentemente da estrutura de redação dos jornais e emissoras de televisão, onde a fonte da informação é conhecida, na Internet pode-se dizer que nem sempre a fonte é conhecida. Isso pode gerar alguns problemas, entre eles, a criação de um ambiente propício para a divulgação de notícias falsas² pode ser instituído, causando um impacto negativo no ecossistema comunicacional.

Ainda, com essa onda de informação circulando nas redes, pode-se pensar em um maior acesso da população ao sistema político, recebendo e enviando diretamente aos

¹ Brexit é a abreviação de Britain Exit que significa "Saída Britânica" e se refere a saída do Reino Unido da União Europeia.

² Sob os problemas relacionados a produção de desinformação há, posteriormente, um capítulo específico que explora o tema com mais propriedade.

representantes de maneira descentralizada, como também, permitindo consultas diretas ao povo por parte destes representantes, como já refletido. Contudo, para alguns, isso pode significar uma erosão da Democracia representativa e o aparecimento de uma espécie de Democracia direta digital.

(...) Na sociedade digitalizada, tudo funciona a partir de mensagens e discursos que acenam com uma rejeição generalizada ao estado de coisas, prometendo soluções e redensões pela descontinuidade e ruptura. A mobilização por meio de redes sociais possibilita vetos e protestos. Mas não o labor argumentativo nem a construção de acordos com base num debate livre e crítico e a consecução de compromissos cívicos. O que é absurdo na democracia representativa, na democracia digital torna-se algo normal ou comum (...) (FARIA, 2019).

A sociedade digitalizada ganha em velocidade e alcance na transmissão da informação, porém a argumentação embasada em elementos factuais do mundo real dilui-se neste oceano virtual da Internet. O poder da influência das manifestações em redes sociais proporciona mudanças que podem não ser as mais acertadas em se tratando de organização política e social, pois o debate sobre o tema em termos de profundidade e pensamento crítico dá lugar a uma superficialidade generalizada. Faria (2019) observa a situação do aparecimento de uma lógica de curto prazo introduzida por essa comunicação em tempo real, criando um cenário de improviso na formulação de políticas públicas.

Pode-se imaginar que o contato direto representante-representados por meio das redes sociais pressione de tal forma que o debate político se torne limitado e de conveniência de modo a não alcançar o objetivo do bem comum em uma sociedade. Não se quer dizer com isso, que essa comunicação seja um inviabilizador da organização política, porém em um Estado democrático é necessário que os objetivos, prioridades e estratégias passem pela negociação política, regras e procedimentos.

Nesse contexto dinâmico de modificação estrutural da esfera pública, a Democracia também sofre influência dessa hiperinformação ventilada nas plataformas de redes sociais. Outros aspectos precisam, também, ser elucidados para que se possa entender do ponto de evolucionário qual a real contribuição que as novas tecnologias para comunicação e integração social.

As plataformas de redes sociais, notadamente, são instrumentos que atualmente dão grande vazão aos debates públicos, não se pode esquecer, porém, que são organizações privadas e que precisam como qualquer outra empresa, apresentar ganhos monetários para se manter ativa. É sob esse ponto de vista que se deve observar a relação do interesse público em um ambiente privado com regras próprias. A transparência é uma variável importante dessa relação e algo que precisa ser de conhecimento geral como as tecnologias tratam as informações que circulam nesses ambientes.

Nesse contexto de grandes mudanças na esfera pública e da infraestrutura de comunicação, é essencial que se apresente algumas informações a respeito do desenvolvimento deste novo modo interação. Dessa forma, são explicitados, a seguir, os principais conceitos e dados sobre algumas plataformas de redes sociais ativas, bem como, suas características e representações econômicas de mercado para que se tenha ideia da força de influência que atualmente possuem.

2.3 A Internet e as plataformas de redes sociais

A Internet sem dúvida proporcionou grandes mudanças na forma em que a sociedade se apresenta, tanto na criação de novos hábitos, como em alterações nas estruturas da economia e políticas. O avanço tecnológico subsidiou novas formas de comunicação que permitem que o fluxo da informação ultrapasse as barreiras territoriais instantaneamente.

A população mundial atingiu em 2021 a marca de 7,9 bilhões de pessoas segundo o sítio de internet especializado em estatística WORLDMETER (2021). A população do Brasil é de 214 milhões de pessoas. Seguindo essa abordagem quantitativa, tem-se ainda que, em outubro de 2021, 4,88 bilhões de pessoas ao redor do mundo tem acesso à Internet, isso significa que aproximadamente 62% da população mundial utiliza a rede mundial de computadores. Ainda, as plataformas de redes sociais possuem um total de 4,5 bilhões de usuários³ (DATAREPORTAL, 2021a).

No Brasil, no início de 2021, havia um total de 160 milhões pessoas com acesso à Internet, o que representa 75% da população nacional. Quando se trata de plataformas de redes sociais, o Brasil possui aproximadamente 150 milhões de usuários⁴ e isso representa mais de 70% da população. Ainda para se ter uma visão mais global do uso da Internet, o brasileiro⁵

³ O número total de usuários nas plataformas de redes sociais pode não significar o total de pessoas, ou seja, uma pessoa pode ter mais de um usuário cadastrado.

⁴ Idem

⁵ Usuários entre 16 e 64 anos de idade.

gasta em média diária 10 horas e 8 minutos de Internet, 4 horas e 2 minutos assistindo televisão⁶, 3 horas e 42 minutos utilizando redes sociais, 3 horas e 13 minutos lendo jornais físicos ou na Internet (DATAREPORTAL, 2021b).

Esses dados são bastante significativos para expressar o potencial de conectividade de usuários na Internet no Brasil. Há uma virtualização das ações cotidianas, o mundo *on-line* se aproxima cada vez mais do mundo *off-line*. Há um tempo significativo sendo gasto nas redes sociais se aproximando bastante da utilização da média de tempo com televisão e já ultrapassando o tempo de leitura de jornais tradicionais.

As plataformas de redes sociais são, atualmente, a principal fonte de informação de 7 em cada 10 brasileiros (BRANCO, 2021). Número expressivo que apresenta uma grande inserção das novas mídias na construção do conhecimento da população nos mais diversos assuntos.

Em uma sociedade cada vez mais conectada e com diversas fontes de informação, é preciso um cuidado no que se consome, pois nem sempre o conteúdo é verdade e o produtor é conhecido, há uma espécie de banalização informacional, em que qualquer um pode escrever sobre qualquer assunto e isso ser disseminado para milhões de pessoas em questão de segundos via plataformas de redes sociais.

Para se conhecer mais sobre essas novas plataformas de informação e sua estrutura básica, é preciso, apresentar alguns conceitos e elementos caracterizadores. Rede social pode ser entendida com um conjunto de dois elementos que formam uma estrutura de grafo⁷: atores e suas conexões. Os atores podem ser pessoas, grupos, empresas, robô, são os nós na rede, já as conexões são os laços sociais que unem esses nós (RECUERO, 2009).

Essas conexões entre os atores de rede podem estar relacionadas aos mais diversos vínculos sociais, como família, localidade, política, saúde, ou seja, algo de interesse comum. Essa rede social pode migrar para Internet, sendo suportada pelas plataformas de rede social. Nesse momento, as possibilidades de conexão entre atores tornam-se virtualmente ilimitadas, sendo possível criar ou acessar redes em todo mundo.

A popularidade de um usuário de uma plataforma de rede social pode ser observada pela quantidade de conexões que ele detém. Trazendo como exemplo a plataforma de rede social *Twitter*, o ex-presidente dos Estados Unidos da América (EUA) possuía 130 milhões de

⁶ Horas gastas vendo televisão via *Broadcast* (acessado por televisão de maneira tradicional) e *Streaming* onde o canal é disponibilizado via Internet.

⁷ A teoria dos grafos ou de grafos é um ramo da matemática que estuda as relações entre os objetos de um determinado conjunto.

seguidores. O que significa que em sua rede social, qualquer informação que seja divulgada por seu usuário, instantaneamente será disponibilizada potencialmente para no mínimo todos estes seus seguidores⁸.

Como já visto, há bilhões de usuários integrando as plataformas de redes sociais, porém é preciso lembrar que são plataformas pertencentes a empresas privadas. Há uma máxima já popularizada na era da informação que diz: “se o serviço é de graça, o produto provavelmente é você”.

Recorrendo, novamente, aos números, a expressividade é mantida quando se avaliam algumas das principais plataformas de redes sociais em termos de quantidade de usuários⁹ ativos. O *Facebook* em outubro de 2021 possuía aproximadamente 2,8 bilhões de usuários, o *YouTube* 2,2 bilhões de usuários, o *Instagram* 1,3 bilhões de usuários e o *TikTok* 1 bilhão de usuários e o *Twitter* 463 milhões de usuários (STATISTA, 2021).

Um marco importante alcançado pelo *Facebook* é que a marca atingiu o valor de US\$ 1 trilhão de valor de mercado em junho de 2021 (ESTADÃO CONTEÚDO, 2021). Fato que só reforça o poderio econômico que as plataformas de rede social possuem e dessa forma a influência que podem se valer para interesses privados.

Com a temática abordada neste capítulo, foi possível observar alguns aspectos da mudança estrutural da esfera pública e a influência das evoluções tecnológicas na produção de novas áreas de debate, ampliando o poderio da circulação da informação.

O debate saiu dos cafés e praças públicas e passou a habitar as plataformas de redes sociais fazendo uso de todo o seu alcance. As informações de interesse público hoje trafegam por meio de ambientes privados, cercados de regras que precisam ser de conhecimento de todos. A transparência é um fator essencial para a saúde dos ambientes virtuais.

O propósito das conexões sociais se amplia e proporciona ao sistema político um canal direto como o povo sem a necessidade de intermediários, o político “fala” aos seus eleitores instantaneamente e da mesma forma o povo retorna ao político seus anseios e desejos. Essa espécie de democracia direta digital vai de encontro a democracia representativa a partir do momento em que o processo político de negociação é esquecido e dá lugar a potencial repercussão momentânea nas redes, a decisão ou voto se fragiliza e é influenciada por volume e não qualidade de análise crítica.

⁸ Nas plataformas de redes sociais, um usuário pode permitir que outros usuários não pertencentes a sua rede de relacionamento acessem suas informações divulgadas em seu perfil. Perfil pode ser entendido como sinônimo de usuário.

⁹ Uma pessoa pode ser usuária de mais de uma rede social, assim o somatório das quantidades não implica o total de usuários em plataformas de redes sociais.

A plataformas de redes sociais vão além de áreas de debate público, o interesse empresarial também se faz presente a partir do momento que a infraestrutura utilizada é privada. O usuário, além de participar das discussões nas redes ou do consumo de qualquer outro tipo de informação, entrega a essas empresas seus hábitos virtuais, seus dados agora registram pegadas virtual que poderão ser usadas para a construção de perfis psicológicos.

O objetivo específico deste capítulo foi apresentar a mudança da esfera pública até a consolidação das plataformas de redes sociais como ambiente transformador do fluxo de informação, que ultrapassa as barreiras territoriais sendo disponibilizada instantaneamente em qualquer lugar do mundo para bilhões de pessoas.

No próximo capítulo são discutidos problemas relacionados a essa nova etapa de intenso fluxo de informação com várias fontes nem sempre confiáveis. Dessa forma, um ambiente propício às campanhas de desinformação em massa é gestado e pode resultar em problemas críticos como a falta de interesse de vacinação em uma pandemia como a da Covid-19¹⁰, nas tentativas silenciar minorias ou mesmo na tentativa de manipulação da população para influenciar resultados de eleições afetando a democracia dos países.

Assim, o próximo capítulo faz uma análise técnica do problema da desinformação em massa apresentado os principais conceitos, classificações e sua estruturação em campanhas nas plataformas de redes sociais, bem como trata da liberdade de expressão no cenário brasileiro e da moderação de conteúdo nas plataformas de redes sociais.

¹⁰ A Covid-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global.

3 LIBERDADE DE EXPRESSÃO, DESINFORMAÇÃO E MODERAÇÃO DE CONTEÚDO EM CONTEXTO

Se comunicar é expor pensamentos e ideias sobre o entendimento da realidade, é passar ao outro a informação que foi construída na intimidade do ser com base em preceitos morais e sociais do conhecimento de mundo de cada indivíduo. A expressão comunicacional pode ser estabelecida de diversas maneiras como, por exemplo, por meio de elementos textuais, gestuais, orais e pictóricos, e assim transpor o limite psíquico e embarcar no mundo dos fatos transformando a realidade dos emissores e receptores da informação.

A Internet, em um sentido já apresentado, possibilitou a evolução da comunicação da atualidade, em que as barreiras de tempo e limites geográficos são quebradas em frações de segundos. A comunicação foi se moldando às novas capacidades de interagir e com isto as formas de expressão ganharam novos patamares. Contudo, essa possibilidade e liberdade de se expressar de maneira instantânea e de longo alcance deve ser garantida, porém não de forma isenta de responsabilidade, pois pode afetar de maneira problemática os direcionamentos da formação e evolução de uma sociedade nesta percepção de hiperinformação.

Nesse contexto, uma nova preocupação surge com a sociedade da informação, pois é preciso observar como a liberdade de expressão, nas plataformas de redes sociais, vem sendo tratada. Situações como a propagação de discursos de ódio, silenciamento de voz das minorias, ataques à democracia, tudo isso tinha um alcance proporcionalmente menor até pouco tempo sem o uso dessas novas formas comunicacionais. Somada a essa situação, há, também, a propagação de campanhas de desinformação, mutando realidades para as quais são convenientes a determinados objetivos de seus idealizadores. Assim, o meio utilizado para propagação assume papel fundamental para o combate aos elementos maléficos desta nova malha informacional. As plataformas de redes sociais se utilizam, muitas vezes, de termo de uso, que são regras a serem seguidas, para a delimitação do que e como o conteúdo pode ser propagado. Contudo, é preciso que se discuta até que ponto uma moderação deste conteúdo se desvincula do aceitável em uma determinada cultura social ou mesmo dos preceitos do Direito.

Neste capítulo é abordada a liberdade de expressão sob a ótica brasileira, bem como conceitua os principais elementos do problema da desinformação em massa, estabelecendo classificações e sua estruturação em campanhas nas plataformas de redes sociais e, ainda, faz uma análise de algumas abordagens utilizadas pelas empresas na moderação de conteúdos de terceiros que disponibilizados em suas plataformas de redes sociais, assim como relaciona os

principais problemas sobre os aspectos da liberdade de expressão atingida por estas moderações.

3.1 A liberdade de expressão no cenário brasileiro

A liberdade de expressão esta voltada à exteriorização de pensamento individual ou de grupo, trazendo ao mundo fático, opiniões sobre os mais variados assuntos, seus conhecimentos e visões de mundo sem eventual interferência ou retaliação do Estado. A Declaração Universal dos Direitos Humanos, em seu artigo 19, conceitua de forma que:

(...)Todo ser humano tem direito à liberdade de opinião e expressão; esse direito inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e ideias por quaisquer meios e independentemente de fronteiras (...)
(ORGANIZAÇÃO DAS NAÇÕES UNIDAS, 1948).

Neste sentido, o professor José Afonso da Silva elenca algumas das formas de exteriorização da expressão, como, por exemplo, pelo exercício das liberdades de comunicação, de expressão intelectual, artística, científica, na recepção e transmissão de conhecimentos (SILVA, 2011).

Como um direito fundamental, a liberdade de expressão tem valor primordial para a constituição de um Estado Democrático de Direito. O debate de ideias é essencial para construção de uma sociedade mais equânime, possibilitando o exercício da soberania popular e trazendo à tona elementos balizadores para a construção de uma democracia sólida e duradoura (CLÈVE, 2022).

O exercício desta liberdade é fundamental para que ocorra a estruturação de um ambiente de fluidez comunicacional entre os mais diversos atores sociais, possibilitando, assim, uma atuação salutar para o enriquecimento dos pilares democráticos. Já a ausência desta liberdade, tem-se um fator limitante para outros direitos e desestabiliza a construção de novas possibilidades de evolução civilizatória.

Este direito fundamental da liberdade de expressão relaciona-se estreitamente com o regime de Democrático de Direito de forma a permitir que os cidadãos possam trabalhar com liberdade e igualdade em busca de uma efetiva participação social e formação da vontade

coletiva, e para isso, devem ter amplo acesso à informação e pontos de vistas dos mais diversos possíveis e assim possam tratar da melhor forma o interesse público (SARMENTO, 2018).

Em um contexto de positivação, a liberdade de expressão está consagrada na Carta Constitucional Brasileira (BRASIL, 2020) e pode ser extraída de seu texto pela observação do art. 5º com os incisos IV, V, IX, XIII em que são assegurados a manifestação de pensamento, o livre exercício de qualquer trabalho, ofício ou profissão, abrangendo também o acesso à informação. Ainda, o art. 220 promove a livre manifestação do pensamento, da criação, da expressão e da informação.

É importante observar que, mesmo dando ampla abertura para a liberdade de expressão, o texto constitucional estabelece alguns dispositivos para que o alcance seja respaldado de legalidade e que seu exercício desvirtuado não esteja acobertado possibilitando danos a terceiros com a invasão em princípios e direitos. Isso fica claro, quando se permite a livre manifestação de pensamento, mas veda-se o anonimato ou quando se assegura o direito de resposta e permite-se a indenização por dano moral ou à imagem.

O caráter não absoluto da liberdade de expressão pode ser sintetizado pela célebre frase atribuída a Oliver Wendell Holmes Jr¹¹ ao dizer que “*a proteção mais rigorosa da liberdade de expressão não protegeria um homem ao gritar falsamente fogo em um teatro e causar pânico*”. Neste aspecto, do ponto de vista do Direito nacional, é possível verificar que há algumas categorias jurídicas que são excluídas de sua proteção, com por exemplo, a calúnia, a injúria, ameaça, incitação ao crime, entre outras tipificadas (FONTENELES, 2021, p. 124).

Um exemplo que figura entre os casos emblemáticos do judiciário brasileiro sobre limitações à liberdade de expressão é o Caso Ellwanger¹², em que o STF, pelo respeito aos princípios da dignidade da pessoa humana e da igualdade jurídica, manteve a condenação de escritor e editor pelo crime de racismo, aplicado pela publicação, venda e distribuição de material antissemita.

Em manifestação de voto no RE 511.961/SP, o Ministro Gilmar Ferreira Mendes esclareceu que as restrições à liberdade de expressão em sede legal são admissíveis, contudo, é preciso que tenham a intenção de promover outros valores e interesses constitucionais também relevantes e respeitando o princípio da proporcionalidade (MENDES, 2017, p.1395).

É possível observar a formação característica da liberdade de expressão no Brasil, pois apesar de não haver uma hierarquia entre os direitos fundamentais, esta possui uma posição preferencial. Esse entendimento tem prevalecido na jurisprudência, concepção esta, já afirmada

¹¹ Oliver Wendell Holmes Jr. foi juiz da Suprema Corte norte americana, advogado e filósofo.

¹² Caso Ellwanger – Julgamento do HC 82.424 no Supremo Tribunal Federal.

pelo ministro do Supremo Tribunal Federal Luís Roberto Barroso em que esclarece que “*A liberdade de expressão desfruta de uma posição preferencial no Estado democrático brasileiro, por ser uma pré-condição para o exercício esclarecido dos demais direitos e liberdades*”¹³.

Essa “posição preferencial” *a priori* foi adotada no julgamento da Arguição de Descumprimento de Preceito Fundamental (ADPF) 130¹⁴ do STF em que se extrai que, na colisão com outros direitos, deve-se usar o princípio da proporcionalidade, partindo-se da premissa que a liberdade de expressão axiologicamente possui um patamar elevado na ordem constitucional brasileira (MENDES, 2017, p.1395).

Situada a importância no texto Constitucional, é interessante observar, dentro de um contexto tecnológico, a valoração da liberdade de expressão como princípio essencial para o uso da Internet no Brasil conforme expressa o art. 8º do Marco Civil da Internet “*A garantia do direito à privacidade e à liberdade de expressão nas comunicações é condição para o pleno exercício do direito de acesso à internet*”. Ainda, neste sentido *caput* o art. 19 promove a seguinte redação:

(...) Art. 19. Com o intuito de assegurar a liberdade de expressão e impedir a censura, o provedor de aplicações de internet somente poderá ser responsabilizado civilmente por danos decorrentes de conteúdo gerado por terceiros se, após ordem judicial específica, não tomar as providências para, no âmbito e nos limites técnicos do seu serviço e dentro do prazo assinalado, tornar indisponível o conteúdo apontado como infringente, ressalvadas as disposições legais em contrário (...) (BRASIL, 2014).

Dessa forma, percebe-se que a liberdade de expressão tem conotação perene e se apresenta de maneira a possibilitar, em seu ponto alto, sustentação à Democracia, promovendo a participação das diferentes vozes da população em busca de um alinhamento social. Ainda, pode-se verificar que é ponto primordial para acompanhar as transformações tecnológicas, como é o caso da Internet.

Cabe, também, apresentar um exemplo do tratamento dado à liberdade de expressão na Internet, como é o caso do Mandado de Segurança Nº 36.666/DF¹⁵ em que um jornalista teve seu usuário do Twitter bloqueado para acesso à conta pessoal do Presidente da República

¹³ STF, 1ª Turma - Reclamação 22.328 Rio de Janeiro, Relatoria do Min. Roberto Barroso

¹⁴ ADPF 130, que derrubou a Lei de Imprensa, do regime militar, e vedou a censura.

¹⁵ Mandado de Segurança Nº 36.666 Distrito Federal

<https://portal.stf.jus.br/processos/detalhe.asp?incidente=5766575>, visualizado em 19 de maio de 2022.

e solicitou desbloqueio pois entendia que há veiculação de informações de interesse social e assim é de interesse público.

A ministra Carmem Lúcia do Supremo Tribunal Federal (STF), relatora do caso, proferiu seu voto com a seguinte passagem:

(...) “a aparente informalidade, suposta precariedade ou privatividade de plataformas digitais, como o Twitter, não tira o caráter oficial das manifestações proferidas pelo presidente da República e que, por isso, o bloqueio de um cidadão e seu afastamento do debate público decorrente de exercício de crítica, ainda que impertinente ou caricata, constitui ato de autoridade pública, adquirindo, nesse contexto, viés censório, inadmissível no ordenamento constitucional vigente” (...).

Ainda, seguindo a mesma linha de raciocínio, o ministro Marco Aurélio de Mello proferiu voto refletindo sobre a censura e liberdade de expressão:

(...) “não cabe, ao presidente da República, avocar o papel de censor de declarações em mídia social, bloqueando o perfil do impetrante, no que revela precedente perigoso”. “Uma vez aberto canal de comunicação, a censura praticada pelo agente político considerada a participação do cidadão, em debate virtual, com base em opinião crítica, viola a proibição de discriminação, o direito de informar-se e a liberdade de expressão, consagrada no artigo 220 da Constituição Federal” (...)

A proteção à liberdade de expressão ultrapassa o mundo físico e vai se inserindo no mundo virtual. Neste cenário, em que as praças públicas estão simuladas em plataformas de redes sociais, é preciso que se tenha atenção para evitar que vozes de minorias sejam silenciadas em seus justos pleitos, mas, por outro lado, também, é importante que se observem as particularidades destas novas ferramentas de comunicação para que, por meio de artifícios tecnológicos, o discurso não seja capturado por narrativas fraudulentas e se consolidem como escolhas sociais viciadas.

Como é possível o observar enquanto elemento garantidor democrático, a liberdade de expressão, ainda que estabelecida em uma posição preferencial, não se sobrepõe ao seu uso indevido como em discursos de ódio, terrorismo, pedofilia, ameaças as instituições democráticas.

Um cenário que merece especial atenção é o da desinformação em massa que, no contexto da Internet, são executados por meio dos ambientes virtuais das plataformas de redes

sociais gerando riscos à liberdade de expressão, assim, a próxima seção apresenta o tema em termos dos principais elementos do problema da desinformação em massa, algumas classificações de desinformação e sua estruturação em campanhas nas plataformas de redes sociais.

3.2 O problema da desinformação em massa

Hoje em dia é muito comum o uso das tecnologias para interações sociais, as facilidades são virtualmente ilimitadas, ao passo que as fronteiras espaciais praticamente deixam de existir. A informação trafega em uma velocidade que transpõe o seu entendimento. E sob esse aspecto são vivenciados cenários de desinformação, onde conteúdo é interpretado ou modificado para alcançar objetivos dos mais diversos e sobretudo em composições de narrativas.

As redes sociais virtuais servem de campo de disseminação em massa de informações muitas vezes duvidosas ou previamente trabalhadas para que atinjam objetivos de distorcer a realidade e sirvam de arma para a obtenção de propósitos econômicos, sociais ou políticos.

Esse desarranjo da informação ocasionando a desinformação não é algo novo, porém ganhou uma roupagem mais efetiva com a ampliação que foi possível de ser obtida com Internet e suas redes sociais. A possibilidade do anonimato, a camuflagem ou mesmo roubo de identidades proporciona ao agente causador um campo amplo de possibilidades para atuar de forma mais livre no alcance de seus objetivos.

O primeiro conceito a ser explorado de desinformação¹⁶ é a construção de Bittman (2019) que esclarece da seguinte forma:

(...) Desinformação é uma mensagem falsa cuidadosamente construída, vazada no sistema de comunicação de um adversário para enganar a elite tomadora de decisões ou o público. A desinformação pode ter natureza política, econômica, militar ou mesmo científica. (...) (BITTMAN, 2019, p. 86)

O termo ainda pode ser categorizado de uma maneira mais ampla quando abordado sob o aspecto de formas de desordens da informação. Nesse sentido, ainda não há uma tradução adequada para o português sendo mais coerente usar o inglês para a explicação.

¹⁶ Um esclarecimento importante ao se tratar do conceito de desinformação é quando se utiliza o termo Fake News, pois este generaliza e é mais utilizado popularmente em jornais para tratar notícias falsas sem levar em consideração as especificidades das desordens da informação, assim para esse trabalho o termo desinformação tem maior coerência científica para tratar do assunto.

Wardle e Derakhshan (2017) fazem uma diferenciação com base na intenção da produção e propagação da informação. Na *disinformation* o conteúdo é falso e criado com objetivo de ser danoso, motivado por ganhos financeiros, influência política, ou simplesmente para causar algum problema a um alvo. A *misinformation* o conteúdo falso é distribuído por pessoas que não sabem dessa falsidade e muitas vezes acreditam que estão ajudando com a divulgação da informação. A *malinformation* são informações verdadeiras divulgadas com o objetivo de causar danos, a exemplo de um vazamento de informações sigilosas.

É importante, também, ter em mente quais são os elementos que movimentam a informação dentro de canais de comunicação. O agente, primeiro elemento, é o produtor ou disseminador da mensagem que pode ou não ter consciência da real motivação. A mensagem propriamente dita e suas características e por fim o intérprete, aquele que recebe e pode produzir reações ao que assimilou.

Outro ponto para entender o fenômeno da desinformação em massa no mundo virtual, é analisar como uma campanha pode ser estruturada, para isto alguns passos são elencados por Schneier (2019):

Passo 1: Encontrar rachaduras no tecido social. Como as divisões sociais, demográficas, econômicas e étnicas;

Passo 2: Espalhar distorções da realidade criando narrativas alternativas;

Passo 3: Envolver essas narrativas em núcleos de verdade. Ou seja, colocar informações falsas no processo de comunicação de informações verdadeiras;

Passo 4: Construir audiências, seja controlando diretamente uma plataforma (como replicar as mensagens no Twitter, por exemplo) ou cultivando relacionamentos com pessoas que serão receptivas a essas narrativas;

Passo 5: Esconder sua responsabilidade; fazer parecer que as histórias vieram de outro lugar;

Passo 6: Cultivar “idiotas úteis” que acreditem e ampliem as narrativas. Incentivá-los a tomar posições ainda mais extremas do que fariam de outra forma;

Passo 7: Negar o envolvimento, mesmo que a verdade seja óbvia;

Passo 8: Jogar pensando no longo prazo. Esforçar-se pelo impacto de longo prazo sobre o impacto imediato.

Como é possível se verificar com os passos apresentados, a desinformação é realizada por meio de uma estrutura lógica pensada para um determinado fim ilegítimo e busca,

para se tornar efetiva, problemas sociais que podem ser explorados. É notório que nas plataformas de redes sociais esse tipo de ação ganha proporção gigantesca e causa um impacto que pode desestabilizar ambientes democráticos, como por exemplo, o caso da *Cambridge Analytica* amplamente divulgado em 2018 com o uso massivo de dados para influenciar o resultado das eleições para a presidência dos EUA (NAPOLI, 2019).

A empresa utilizou grandes quantidades de dados de mídia social para construir perfis psicológicos, demográficos e geográficos detalhados de eleitores. Após os perfis criados foram enviadas mensagens de cunho político por meio de plataformas de rede social como o Facebook manipulando a informação para que o resultado se desse de acordo com contratado.

Um problema a ser enfrentado com a desinformação, onde as plataformas de redes sociais são o *locus* da circulação da informação, é que na medida que se passa a combater esse tipo de ação, pode haver repercussão no uso legítimo da liberdade de expressão. Assim, mecanismos de moderação de conteúdo digital devem procurar encontrar limites entre o que é legítimo e o que é abuso de direito de se expressar, por exemplo. Na seção seguinte é feita uma análise de algumas abordagens utilizadas pelas empresas (plataforma de redes sociais) na moderação de conteúdos de terceiros que são disponibilizados em suas plataformas de redes sociais.

3.3 Moderação de conteúdo em plataformas de redes sociais

O fluxo da malha comunicacional das plataformas de redes sociais é, como já percebido, demasiadamente dinâmico e a informação se espalha aos quatro cantos do mundo instantaneamente. O conteúdo dessa informação em alguns casos pode violar regras de conduta postas pela plataforma, regras sociais, bem como, do ordenamento jurídico nacional. Dessa forma, há uma constante evolução de como a remoção do conteúdo dessas plataformas é realizado. Entende-se como moderação de conteúdo o conjunto de mecanismos de governança que estruturam a participação em uma plataforma para facilitar cooperação e prevenir abusos (MARANHÃO et al., 2020).

Neste sentido, Klonick (2018) descreve em seu artigo “*The New Governors: The People, Rules, And Processes Governing Online Speech*” um breve relato de como a moderação de conteúdo foi se desenvolvendo ao longo dos anos, e serve como base para a explanação nesta seção. Inicialmente, em 2009, no Facebook, toda a moderação de conteúdo era baseada em algumas “regras” internas aplicadas globalmente a todos os usuários. Dessa forma, fotos de pessoas nuas, apologia ao nazismo entre outras informações estavam em uma lista de

moderação e eram removidas da plataforma. A lista não apresentava distinção de conteúdo ou o porquê de a informação constar dentro dela.

Posteriormente, houve uma categorização distinguindo entre elementos para moderação, por exemplo, a diferenciação entre arte e nudismo, discursos de ódio, entre outros. Em seguida, foi evoluindo de um sistema inicial de padrões para um sistema de regras. Essa evolução se deu principalmente pela observação do rápido aumento de usuários e volume de conteúdo, a globalização e a diversidade de comunidades online e a crescente dependência de equipes de moderadores humanos com diversas formações.

Segundo a autora, a moderação de conteúdo acontece em diversos níveis, de maneira que ocorre antes da publicação na plataforma, que pode ser identificada como *ex ante*, ou posteriormente a publicação do conteúdo, como na moderação *ex post*. Ainda, podem ser realizados de maneira reativa, isto é, após os usuários da plataforma de rede social chamarem atenção para o conteúdo, ou de maneira proativa em que os moderadores realizam ativamente verificações de conteúdo passíveis de remoção.

A moderação humana de conteúdo nas plataformas atua semelhante à de um juiz, pois os moderadores são treinados para exercer julgamento profissional sobre a aplicação das regras internas de uma plataforma e, ao aplicar essas regras, espera-se que os moderadores usem conceitos legais como relevância, razão por meio do exemplo e analogia.

Além destas formas de moderação, há uma diferenciação entre a automatizada e a manual. Na automatizada o algoritmo recebe, por exemplo, denúncias de conteúdo inapropriado e pode retirar a informação da plataforma (GRIMMELMANN, 2015). As formas de moderação de conteúdo também podem ser realizadas para determinadas áreas geográficas, assim, um conteúdo considerado irregular para aquele país, por exemplo, pode ser excluído, neste caso é preciso ter cuidado com diferenciação entre o lícito e ilícito de acordo com a localidade.

Segundo Maranhão et al. (2020) há algumas preocupações relevantes sobre o impacto em direitos fundamentais e elenca algumas ordens de questionamentos em relação aos moderadores:

(...)A primeira diz respeito à adequação dos padrões de moderação, tipicamente modulados para democracias constitucionais, assumindo-se, por exemplo, que a desinformação e o ódio são criados e tolerados apenas por uma minoria, quando na verdade podem ter o suporte da maioria da população e ter inclusive origem governamental. A segunda concerne à necessidade de competência dos moderadores para lidar com questões locais, incluindo o conhecimento da língua e aspectos culturais e políticos relevantes, considerando que, por vezes, a moderação não é nacional e até mesmo pode ser executada de modo terceirizado para empresas

estrangeiras(...). (...) A terceira preocupação concerne às condições de trabalho do moderador, que é exposto constantemente a imagens e conteúdo degradantes, o que pode afetar sua saúde psíquica (...).

Já para a automação da moderação, há críticas atribuídas ao excesso de confiança no potencial de acurácia das ferramentas podendo haver violações à liberdade de expressão, pelo excesso de bloqueios entre outros problemas relacionados à direitos humanos. Nesse sentido é necessário que haja uma transparência de como essas atividades são realizadas.

É interessante, também, apresentar alguns dados que são disponibilizados pelas plataformas quanto a estatística de moderação de conteúdo realizado. Por exemplo, o Twitter em janeiro a junho de 2021 teve 4.8 milhões de contas acionadas¹⁷, 1.2 milhões de contas suspensas¹⁸ e 5.9 milhões de conteúdo removidos¹⁹ (TWITTER, 2022). Já o Youtube, teve de outubro a dezembro de 2021 3.850.275 canais removidos por alguma violação das diretrizes estabelecidas nos termos de uso. Destes 89.8% foram relativos spam, conteúdo enganoso e golpes (GOOGLE, 2022a).

Interessante, também, observar que algumas vezes o conteúdo não é removido, pois, informa o Google, que há solicitações, de governos, por exemplo, não são específicas o suficiente para saber o que deve ser removido, com descrições genéricas do tipo remover da “Internet”. Em outras situações, o próprio usuário remove o conteúdo antes não havendo mais necessidade (GOOGLE, 2022b).

A moderação de conteúdo é algo que precisa ser bem estruturado para que trate da determinas regras e seja responsivo o suficiente para evitar que o tráfego de informações reprováveis legalmente e socialmente, porém, de fato há uma dificuldade em estabelecer estes limites.

Uma informação recente sobre moderação de conteúdo no Facebook apresentada pela reportagem investigativa do *Wall Street Journal* (WSJ)²⁰ repercutiu bastante nas discussões sobre o tema. A investigação se valeu de revisões de documentos internos da empresa investigada, incluindo relatórios de pesquisa, entrevistas com empregados e rascunhos de apresentações da alta administração.

¹⁷ As “Contas acionadas” refletem o número de contas únicas que foram suspensas ou tiveram algum conteúdo removido por violar as Regras do Twitter.

¹⁸ As "Contas suspensas" reflete o número de contas únicas que foram suspensas por violar as regras do Twitter.

¹⁹ O “Conteúdo removido” reflete o número de partes exclusivas de conteúdo (como Tweets ou imagem de perfil, banner ou biografia de uma conta) que o Twitter exigiu que os proprietários de contas removessem por violar as Regras do Twitter.

²⁰ Matéria do Wall Street Journal – the facebook files. Disponível em: <https://www.wsj.com/articles/the-facebook-files-11631713039>. Acesso em: 10 de maio de 2022.

Aparentemente uma série de problemas identificados que geraram efeitos nocivos não haviam sido corrigidos, mesmo sendo tema de repetidas audiências no congresso norte americano e denúncias da mídia. Um dos achados que afeta diretamente à liberdade de expressão é que a empresa afirma que todos os usuários estão igualmente sujeitos às regras da plataforma para publicação de conteúdo. Porém, a empresa, de maneira não transparente, segundo a reportagem, construiu um sistema²¹ que isentou perfis, considerados pela por ela de alta importância, de algumas ou senão todas as regras de moderação. A reportagem relata que muitos desses perfis abusaram do privilégio, publicando material incluindo assédio e incitação à violência que normalmente levaria a sanções.

Um exemplo desta situação, segundo a reportagem do WSJ, foi o caso do jogador Neymar Jr. que figurava na lista dos VIPs para a empresa, assim com privilégios quanto a moderação do conteúdo. O jogador respondeu a uma acusação de estupro em 2019 postando vídeos no Facebook e no Instagram se defendendo, e para isso publicou conversas WhatsApp com o nome da denunciante e fotos dela nua. Em vez de excluir imediatamente o material, que é o procedimento do Facebook para “imagens íntimas não consensuais”, os moderadores foram impedidos por mais 24 horas de remover o vídeo. Assim houve a visualização do conteúdo aproximadamente 56 milhões de vezes. Nos próprios documentos internos, as postagens foram consideradas como “Vingança pornográfica”.

Esta situação pode ser analisada a luz do princípio da igualdade, abarcado na Constituição Federal, que em seu *caput* do art. 5º (BRASIL, 2020) onde esclarece que todos são iguais perante a lei, sem distinção de qualquer natureza garantindo-se o direito à igualdade. Neste princípio são vedadas as diferenciações absurdas e arbitrárias como parece ser o caso em questão, que pelo fato de tratar desigualmente os usuários da rede, permitindo que alguns tenha uma espécie de imunidade de moderação traz à tona uma afronta a tal princípio e assim criando mais um problema a ser tratado pelos meios legais.

Assim, como apresentada nesta seção, percebe-se que o tema da moderação de conteúdo se mostra bastante complexo, levando em consideração o poderio das empresas de plataformas de redes sociais na esfera pública e seus espaços para decisões. De fato, é necessário também um olhar Estatal, não no sentido estrito, mas em conjunto com a área privada e academia, encontrem formas de se prevenir a disseminação de conteúdo fraudulento, ou mesmo que cause impactos nas liberdades, sobretudo na liberdade de expressão.

²¹ O programa, conhecido como “cross check” ou “XCheck”, foi concebido como uma medida de controle de qualidade para contas de VIP, contudo acabou proteger milhões de VIPs da fiscalização normal da empresa.

No próximo capítulo, são apresentadas em linhas gerais a experiência alemã na regulação das plataformas de redes sociais e o instituto autorregulação regulada. A iniciativa do Projeto de Lei 2630/2020 que institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet também é apresentado em seus principais pontos no contexto deste trabalho com algumas considerações sobre a proteção à liberdade de expressão.

4 REFLEXÕES SOBRE A REGULAÇÃO DAS PLATAFORMAS DE REDES SOCIAIS

Como visto nos capítulos anteriores, nos últimos anos houve uma revolução nos meios de comunicação, alcançando patamares próximos aos pensados apenas em histórias de ficção científica. A Internet proporcionou mudanças econômicas e sociais e novas aplicações foram surgindo, a exemplo das plataformas de redes sociais. A conexão entre pessoas, hoje em dia, desconhece barreiras geográficas e a mensagem entre o emissor e receptor é instantânea.

Alguns problemas foram surgindo com a massificação do uso das plataformas de redes sociais, como, por exemplo, a propagação de discursos de ódio, campanhas de desinformação, ataques à Democracia, entre outros. Isso intensificou a discussão a respeito da responsabilização das plataformas sobre o conteúdo veiculado em suas redes sociais.

No contexto legal, a trajetória para este debate em termos de moderação de conteúdo pode ser observada com o *Communications Decency Act (CDA) de 1996* nos Estados Unidos que em sua seção 230 afirma que “*No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.*”. Contudo, não é um salvo-conduto para inexistir a possibilidade de remoção de conteúdo, pois pode haver, por exemplo, a remoção de material ilegal em nível de lei federal daquele país (WIKIPEDIA, 2022).

No Brasil, o Marco Civil da Internet em seu art. 19 também se manifesta sobre o tema, assim, há responsabilização de plataformas de redes sociais se estas não tomarem providências para remover conteúdo após recebimento de ordem judicial, por exemplo. Ainda, a ordem judicial tem apenas a intenção de responsabilização posterior da plataforma (responsabilização do intermediário), o que não impede que a própria plataforma tome a iniciativa, pois está livre para implementar as políticas que entendem ser pertinentes para remoção voluntária de conteúdo (STROPPA, et al., 2022).

A moderação de conteúdo é um tema que precisa ser analisado sob diversas óticas para que algumas liberdades não sejam afetadas, principalmente a liberdade de expressão, algo que, atualmente é exercido por meio de ferramentas que potencializam o poder do discurso pelo alcance que é possível se conseguir em ambiente virtual.

Assim, neste capítulo, são apresentadas iniciativas sobre regulação das plataformas de redes sociais em especial no tratamento dado ao conteúdo de terceiros abrigados nesses novos sistemas de comunicação em massa. Inicialmente é feita uma exposição da experiência alemã com a lei para melhoria da aplicação da lei nas redes sociais a NetzDG

(*Netzwerkdurchsetzungsgesetz*)²² e que pode-se considerar referência sobre o tema. Além disto, é apresentada, neste contexto, o instituto da autorregulação regulada como meio de atuação proposto para moderação de conteúdo para caso complicados e a ilegalidade não é de fácil definição.

Em seguida, é abordado o projeto de lei 2630 de 2020, iniciativa brasileira que tem por objetivo instituir a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet e atualmente é tema de diversos debates e tem recebido diversas emendas complementares. Neste trabalho, apenas o texto original é visitado.

4.1 A experiência alemã e a autorregulação regulada

A Alemanha é pioneira na legislação para o tratamento dessa nova realidade digital das plataformas de redes sociais. Pois, tendo observado o *Brexit*²³ e eleições américas, a classe política do país deu uma maior atenção ao tema em função da influência que, hoje em dia, essas plataformas conseguem exercer na esfera pública (CAMPOS, 2017).

Nesse contexto, foi elaborada a *Netzdurchsetzungsgesetz* ou NetzDG, lei que entrou em vigor em 2017 e visa melhorar o ambiente nas redes sociais. Atua de maneira fundamentada em um mecanismo de sanção, como necessidades de moderação de conteúdo em prazos estabelecidos e multa pelo não cumprimento e um mecanismo de *compliance* para tratar as questões de regulação das plataformas e responsabilizações com a apresentação de relatórios e procedimentos formalizados. Ao longo dessa seção são apresentadas algumas considerações sobre a lei.

Primeiramente, o âmbito de sua aplicação são os provedores de serviços de telecomunicação que visem lucro e possibilitem o compartilhamento de conteúdos indeterminados entre usuários ou que possibilite a disponibilização ao grande público destes conteúdos e devem possuir mais de 2 milhões de usuários registrados no país (Ex: Twitter, Facebook, Instagram). Não são alcançados pela lei, as plataformas de cunho jornalísticos e redacionais, nem aquelas que são direcionadas as comunicações individuais (Ex: Whatsapp, Telegram).

A lei apresenta em seu Artigo 1, §1, (3) os conteúdos ilícitos considerados, e estes já são tipificados no Código Penal alemão. Entre estes ilícitos, estão a incitação à prática de

²² A NetzDG. Disponível em: https://www.bmj.de/DE/Themen/FokusThemen/NetzDG/NetzDG_EN_node.html. Acesso em: 05 de jun. de 2022.

²³ *Brexit* é uma abreviação para "*British exit*", ou seja, a saída britânica da União Europeia.

crime grave contra o Estado, incitação ao ódio, disseminação de imagens de violência, ameaça à prática de crime, insulto e difamação. A proteção à direitos autorias ou proteção aos dados pessoais não são referenciados pela NetzDG (CUEVA, 2020).

A transparência é um fator crucial exigido pela NetzDG (Artigo 1, §2), assim, as plataformas de redes sociais, caso tenham recebido mais de 100 reclamações de conteúdos ilícitos, devem produzir relatórios sobre os procedimentos e medidas adotados. Entre os apontamentos que devem figurar nesse relatório estão: o dever de informar quais esforços empreendidos para impedir ações ilícitas na plataforma; quais os mecanismos que estão disponibilizados para formalizar as reclamações sobre conteúdos ilícitos; quais os critérios para a tomada de decisão do apagamento ou bloqueio do conteúdo. Faz parte também deste relatório, a divulgação informações sobre treinamentos realizados pelos recursos humanos que são responsáveis pelo tratamento das reclamações.

O tratamento das reclamações para conteúdos ilícitos deve possuir um procedimento contínuo, efetivo e transparente e estas informações devem estar disponíveis aos usuários de forma fácil e sem interrupções de acesso. Mais uma vez é percebido que é fundamental o fornecimento de transparência ao usuário nas ações realizadas pelo provedor de na moderação de conteúdo.

A lei prevê espécies de gatilhos que iniciam o procedimento de remoção ou bloqueio de conteúdo, assim, situações em que o conteúdo se demonstre evidentemente ilícito o provedor de rede social tem o prazo de 24 horas após o recebimento da reclamação para que as ações sejam efetivadas e o conteúdo removido ou bloqueado. Ainda, caso o conteúdo ilícito necessite de mais análise para se determinar a sua ilegalidade tem-se um prazo padrão de até sete dias. Quando o conteúdo não é tão claro quanto a ilicitude pelas circunstâncias que se apresentam, existe a possibilidade do direito de defesa antes da tomada de decisão, neste caso, o prazo padrão pode ser ultrapassado.

O provedor de rede social pode acionar a instituição de autorregulação regulada para tomada de decisão em relação ao conteúdo analisado. A autorregulação regulada, é apresentada posteriormente nesta seção, contudo resumidamente, trata-se de um instituto independente acreditado pelo Estado para conduzir a análise de conteúdos mais difíceis de se ter uma definição sobre sua ilicitude.

Para efeitos da lei, a autorregulação regulada deve ser composta por diversos provedores de redes sociais ou instituições que detenham estrutura e equipamentos adequados para o trabalho, deve ser independente para ser mais consistente a imparcialidade, deve possuir equipe com conhecimentos técnicos garantidos, deve estabelecer procedimentos claros que

regulem a abrangência e a análise do conteúdo de forma tempestiva e garantir a revisão das decisões se necessário.

Outro ponto importante é a previsão de multas pecuniárias de até 50 milhões de Euros no caso de provedor de rede social descumprir a lei. Por exemplo, caso o conteúdo evidentemente ilícito não seja retirado no prazo de até 24 horas após o recebimento da denúncia está passível de multado.

Essa legislação é pioneira na regulação de plataformas digitais e possui algumas críticas que podem servir de insumo para análise na implementação deste tema em outros países, principalmente no Brasil com andamento o PL2630/2020.

Sob ótica alemã, a primeira crítica a ser observada é sobre sua aventada inconstitucionalidade material pelo argumento de que há uma intervenção Estatal desproporcional e que atinge as liberdades e assim criaram-se ameaças ou processo comunicacional (CUEVA, 2020).

(...)Aponta-se a inconstitucionalidade formal da NetzDG, tendo em vista que a matéria disciplinada é de competência dos Länder²⁴. Haveria também inconstitucionalidade material, caracterizada por uma desproporcional intervenção do Estado nas liberdades protegidas por direitos fundamentais dos provedores de redes sociais, que ameaçam o processo de comunicação(...) (CUEVA, 2020, p. 276-277)

Contudo há a visão que a NetzDG, estabelece com as suas diretrizes, meios para tomada estruturada de responsabilidade de intermediários (provedores de redes sociais) e assim, promove de forma fundamental o caminho correto para a regulação e auxilia na consolidação da liberdade de expressão online (EIFERT, 2020).

Ainda, na visão de CUEVA (2020) as principais críticas à NetzDG estão voltadas para a suposta restrição liberdade de expressão e liberdade de imprensa. Nesse caso, argumenta-se que o Estado transfere à entes privados uma espécie de censura prévia. Para contrapor a esta argumentação, esclarece que a lei prevê uma serie de exigências (já apresentadas na seção) para estabelecer transparência sob os procedimentos adotados na moderação de conteúdo e para que sejam elaborados de relatórios consolidando as informações de moderação, o que em tese já é capaz de uma maior segurança para a realização das ações. Ainda, a autorregulação regulada

²⁴ Competência do Estadual

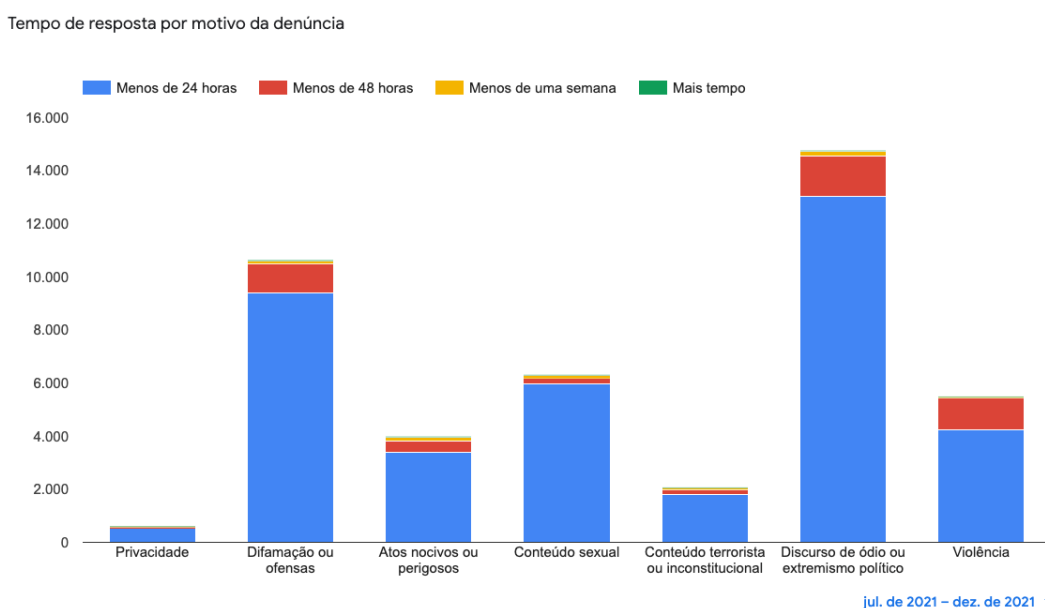
ao congregar empresas provedoras de redes sociais fortalece o controle mútuo quanto aos problemas apresentados.

Outra situação que foi fonte de crítica em relação à restrição da liberdade de expressão é que a lei poderia estar sistematicamente fadada a, em caso de dúvidas, principalmente em alguns conteúdos difíceis em ter sua ilicitude identificada, retirar conteúdos lícitos como efeito colateral. Isso, ocasionaria em *overblocking*, ou seja, conteúdos legais ou de ilicitude duvidosa seriam moderados de maneira a evitar problemas de *compliance* com a lei ou multa, e assim afetaria a liberdade de expressão (EIFERT, 2020).

Contudo, passados alguns anos desde a entrada em vigor da NetzDG, Eifert (2022) tece alguns comentários da experiência da execução da lei. Relata que, por meio de relatórios disponibilizados pelas plataformas, foi possível observar que houve uma melhoria significativa no gerenciamento do tratamento das reclamações e da moderação de conteúdo ilegal, ocasionadas pelo aumento de transparência dos procedimentos adotados.

A figura 1 apresenta um gráfico retirado do relatório de transparência em função da aplicação da NetzDG. Neste caso, é registrado a quantidade categorizada por tipos de denúncias e seu tempo de resposta de atuação em um espaço amostral do período de julho de 2021 até dezembro de 2021.

Figura 1 – Exemplo de gráfico apresentado em relatório de remoção ou bloqueio de conteúdo do Youtube



Fonte: Relatório de Transparência - Remoções de acordo com a Lei aplicável às redes²⁵

²⁵ Um exemplo de relatório sobre remoção de conteúdo do Youtube de acordo com a Netzdg pode ser encontrado no portal de transparência do Google. Disponível em: https://transparencyreport.google.com/netzdg/youtube?hl=pt_BR. Acesso em: 3 de junho de 2022

Outro fator importante ressaltado em sua análise, é que não foram relatadas ocorrências de bloqueios excessivos pelas medidas impostas pela lei, o que poderia acarretar *overblocking*. Ainda, verificou-se que houve muito mais exclusões de conteúdo devido a violações dos termos de uso do contrato da plataforma de rede social (que são condutas definidas e que o usuário precisa concordar para poder participar da rede social) que por força da NetzDG.

Ainda segundo o relato, a lei passou por uma revisão e assim algumas situações deverão ser aprimoradas como, por exemplo, os relatórios de transparência que serão elaborados para permitir comparações entre os provedores, e deverá, também, ser aumentada a clareza dos canais de reclamações a fim de evitar possíveis excessos de bloqueio. Dessa forma, há a promoção de um reforço aos direitos dos usuários para que seja permitida a solicitação de revisão das decisões de moderação dos provedores da rede social, com a inclusão de órgãos de arbitragem imparciais.

Cabe, ainda, enfatizar um dos pilares da NetzDG que é o instituto da autorregulação regulada onde é possibilitado aos provedores de redes sociais transferirem a decisão sobre a ilegalidade de conteúdo para uma instituição acreditada pelo Estado²⁶. O primeiro órgão desse tipo foi reconhecido em janeiro de 2020 já tendo tratado de mais de 14.000 requisições. A seguir é apresentada uma visão geral sobre a autorregulação regulada.

4.1.1 A autorregulação regulada

A velocidade da informação em circulação nas redes sociais é de fato algo surpreendente e ao mesmo um desafio dos tempos modernos, pois o conteúdo é bastante abrangente e pode algumas vezes possuir características de ilegalidade. Essa dinâmica do mundo digital traz uma necessidade de celeridade de atuação para que não se fuja de um contexto democrático e gere problemas sociais em larga escala. Nesse cenário de incerteza, o legislador precisa lançar vistas às possibilidades modernas e criativas de se compor mecanismos de organização do mundo digital.

O instituto da autorregulação regulada é um exemplo de como pode ser tratada essa dinamicidade e incertezas deste momento de intenso uso das plataformas de redes sociais para interação comunicacional. A autorregulação regulada tenta conciliar duas abordagens conhecidas: a autorregulação e regulação (MARANHÃO; CAMPOS, 2022).

²⁶ A primeira instituição de autorregulação regulada reconhecida pelo Estado é a FSM – *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter*, <https://www.fsm.de/en/fsm/netzdg/> acesso em 28 de maio 2022

A autorregulação traz uma eficiência pelo conhecimento interno da indústria e constante revisão de conceitos, que o Estado, por si só, não consegue acompanhar por suas estruturas tradicionais e por não ter o conhecimento técnico para essas novas demandas. Contudo há, também, de se observar que o interesse privado não necessariamente possui os mesmos valores e interesses públicos, sendo esta uma desvantagem. Com a regulação tem-se a possibilidade do exercício do poder coercitivo do Estado atuando de forma a vincular ao interesse público, porém não tem a velocidade de adaptação que se faz necessária para este cenário dinâmico e de incerteza das redes sociais.

A autorregulação regulada surge, nesse contexto das plataformas de redes sociais, de modo a aproveitar as vantagens da regulação e da autorregulação, assim conceitualmente pode-se chegar ao seguinte entendimento:

(...)A autorregulação regulada oferece outra e nova possibilidade de lidar com as incertezas, ao conciliar vantagens das duas abordagens alternativas: a regulação versus a autorregulação. Foca-se no importante momento de auto-organização conforme expertise e dinâmica própria da indústria, estimulando-se, porém, alguns parâmetros gerais de interesses públicos caros ao Estado e sociedade(...) (MARANHÃO; CAMPOS, 2022)

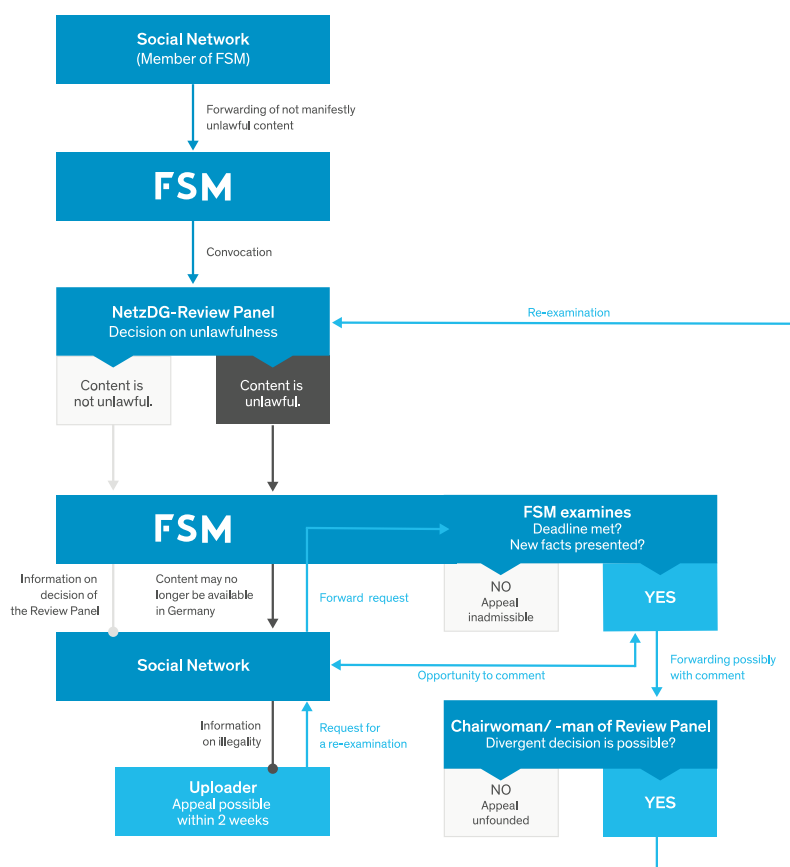
Essa divisão de trabalho entre Estado regulador e os atores sociais a serem regulados em uma situação de cooperação é vantajosa para a sociedade da informação e o cerne da autorregulação regulada pode ser descrita a seguir:

(...)num contexto de crise da regulação estatal tradicional pelo aumento da complexidade social atrelado à absorção por atividades privadas de tarefas estatais, a autorregulação regulada visa conjugar ambas as dimensões da sociedade. Isto é, objetivos públicos orientados pelo interesse público ligado ao conhecimento setorial da sociedade privada em implementar tais objetivos(...) (CAMPOS; GRINGS, 2022)

Em Campos e Grings (2022) esclarecem que a correção ou autorregulação regulada pode ser entendida como uma estrutura que combina elementos da autorregulação, advindas da autonomia privada de pessoas singulares ou coletivas, e da heterorregulação, que se apresenta como espécie de intervenção externa aos sujeitos regulados, usualmente para persecução de determinados interesses públicos.

A NetzDG institui o uso da autorregulação regulada, assim as plataformas podem transferir decisões sobre moderação de conteúdo supostamente ilegal postados em suas redes sociais para uma instituição independente, acreditada pelo Estado (§ 3 § 2 n.º. 3 b, parágrafos 6-8), que possua recursos humanos com conhecimento técnico, procedimentos transparentes e infraestrutura adequada e essa análise tem o prazo de sete dias, sendo possível também a revisão de suas decisões caso necessária.

Figura 2 - Fluxo da Autorregulação Regulada NetzDG - FSM



Fonte: FSM – *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter*²⁷

A figura 2 traz um exemplo do funcionamento da moderação de conteúdo em uma instituição de autorregulação regulada acreditada pelo Estado alemão. Em linhas gerais a execução começa quando uma rede social membro tem um caso de conteúdo que se apresenta de difícil identificação sobre sua ilegalidade, e assim precisa de uma análise independente para saber qual a ação a ser adotada. Dessa forma, é enviada para o instituto, neste caso o FSM. O

²⁷ Fluxograma do funcionamento da moderação de conteúdo em uma instituição de autorregulação regulada conforme a NetzDG, nesse caso a primeira instituição acreditada pelo Estado Alemão, a FSM. Disponível em: <https://www.fsm.de/en/fsm/netzdg/>. Acesso em 29 de maio de 2022.

FSM por sua vez forma um comitê de revisão com integrantes capacitados e independentes para que façam a análise. Se o conteúdo é considerado legal, a rede social é informada e se encerra o fluxo. Contudo, no caso de o conteúdo seja considerado ilegal, a rede social é informada e expõe ao usuário o porquê da moderação, apagamento ou bloqueio. Caso o usuário não concorde, tem duas semanas para requerer uma revisão do resultado da moderação, assim envia a solicitação à rede social que encaminha para análise da FSM, neste momento é verificado se há elementos novos para análise e se o prazo de reconsideração foi respeitado, caso haja algum impedimento o recurso não é admitido. Não havendo impeditivos, a rede social pode ser chamada para realizar algum comentário, e análise é passada para o presidente do comitê de revisão que pode inadmitir a contestação ou passar para nova análise do comitê de revisão, caso entenda que há possibilidade de mudança de resultado, e assim segue o fluxo novamente.

O instituto da autorregulação regulada é utilizado, também, com previsão na lei, em órgãos de arbitragem para facilitar a solução extrajudicial de disputas entre usuários afetados com a moderação do conteúdo e as plataformas de rede social. São instituições organizadas sob o regime do direito privado, reconhecidas pelo Estado, com os requisitos obrigatórios de garantia de qualidade orientados pelo interesse público (EIFERT, 2022).

O Projeto de Lei 2630/2020, tema da próxima seção, também tem a previsão para o uso da autorregulação regulada. A experiência desse instituto em execução na Alemanha pode trazer elementos para o embasamento fático observados nas lições aprendidas e assim, no Brasil, se trabalhe em uma legislação coerente e próxima o suficiente de suprir os desafios da sociedade cada vez mais digital.

4.2 A iniciativa do PL2630/2020

O cenário de desinformação, hoje nas plataformas de redes sociais, tem levado países a discutirem ideias de como combater esse problema antigo, mas com roupagem moderna e alcance instantâneo para milhões de pessoas. Foi apresentada, na seção anterior, a iniciativa alemã, já em execução, que busca, por meio de uma regulação da moderação de conteúdo nas redes sociais, atuar de forma transparente no a combate à desinformação. Neste sentido, os atores políticos brasileiros começaram uma discussão sobre tema, e foi proposto o projeto de lei 2630 de 2020²⁸ de autoria do senador Alessandro Vieira que trata da instituição de Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet.

²⁸ Projeto de Lei 2630/2020 e seus apensados. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2256735>. Acesso em: 29 de maio de 2022.

Estruturalmente²⁹, o PL2630/2020 possui 36 artigos distribuídos em sete capítulos e tem a intenção de criar medidas para combater as a desinformação nas plataformas de redes sociais e nos serviços de mensageria³⁰, não estando dentro do projeto a intenção regular o uso de serviços de e-mail. O art. 3º apresenta os princípios em que a lei tem sua base, e para este trabalho destacam-se: I – liberdade de expressão e de imprensa; IV – responsabilidade compartilhada pela preservação de uma esfera pública livre, plural, diversa e democrática; V – garantia da confiabilidade e da integridade dos sistemas informacionais; VII – acesso amplo e universal aos meios de comunicação e à informação.

O texto procura se alinhar com a Carta Constitucional trazendo como princípios os elementos de garantia de liberdade e responsabilidade na atuação dos provedores de redes sociais. Assim, se reconhece uma opção pelo respeito aos art. 5º incisos IV, V, IX, XIV e art. 220 da CF (BRASIL, 2020).

O projeto de lei consolida, também, alguns objetivos a serem alcançados que são essenciais para o tratamento e combate à desinformação, bem como foca na garantia da liberdade de expressão online, assim se apresenta no art. 4º:

(...)Art. 4º Esta Lei tem como objetivos:

I – o fortalecimento do processo democrático por meio do combate ao comportamento inautêntico e às redes de distribuição artificial de conteúdo e do fomento ao acesso à diversidade de informações na internet no Brasil;

II – a defesa da liberdade de expressão e o impedimento da censura no ambiente online;

III – a busca por maior transparência das práticas de moderação de conteúdos postados por terceiros em redes sociais, com a garantia do contraditório e da ampla defesa;

e IV – a adoção de mecanismos e ferramentas de informação sobre conteúdos impulsionados e publicitários disponibilizados para o usuário(...)

Alguns pontos são destacados no projeto, como a proibição de contas falsas criadas para simular a identidade de terceiros, contudo, há exceção de conta criada e identificada para conteúdo humorístico e paródia, semelhante ao que ocorre na NetzDG. O uso de robôs também é observado para que os provedores de redes sociais e serviços de mensageria criem políticas de uso permitido, nesse caso robôs para uso lícito, e que sejam criados mecanismos de

²⁹ Reportagem da Agência de Notícia da Câmara que resume os principais pontos do PL2630/2020. Disponível em: <https://www.camara.leg.br/noticias/673694-projeto-do-senado-de-combate-a-noticias-falsas-chega-a-camara/>. Acesso em: 29 de maio de 2022.

³⁰ São considerados serviços de mensageria, aplicativos de comunicação como WhatsApp, Telegram, Signal entre outros e são diferentes de redes sociais.

identificação de contas que possuem atuação atípica. Extrai-se que ideia é que se crie uma barreira para dificultar os disparos em massa e orquestração de campanhas de desinformação.

Há também previsão para que seja limitado o número de vezes que uma mesma mensagem é enviada aos usuários e grupos, assim, um conteúdo falso poderia ter seu alcance diminuído e o impacto reduzido, por outro lado, poderia se entender como um limitador da liberdade de expressão. As empresas deverão, também, guardar pelo prazo de três meses registros dos envios de mensagem em massa, o que a depender das informações a serem guardadas pode haver um custo alto para o armazenamento. A publicidade veiculada nas redes sociais deverá ter identificação dos responsáveis para que seja permitido o contato com o anunciante, neste caso se promove uma maior transparência. As contas em redes sociais do presidente da república, governadores, prefeitos, ministros de Estado, parlamentares, entre outros agentes políticos serão consideradas de interesse público e assim o acesso por terceiros às informações publicadas não poderá ser restrito, o que evitaria os casos relatados na seção que foi desenvolvido conceitos de liberdade de expressão.

O PL cria o Conselho de Transparência e Responsabilidade na Internet que tem como função a realização de estudos, pareceres e recomendações sobre liberdade, responsabilidade e transparência na internet e é vinculado ao Congresso Nacional, sendo composto por 21 membros com representantes do poder público, do setor privado, da sociedade e da academia. Cada membro do Conselho terá um mandato de 2 anos, permitida uma recondução.

As empresas que se inserem neste PL deverão ter sede no Brasil e nomear representantes legais, assim os contatos com os órgãos de Estado fica mais célere. Devem produzir e divulgar relatórios de transparência sobre o atendimento às ordens da Justiça brasileira e, além disso, devem manter o registro de dados referentes aos usuários e conteúdos para facilitar o cumprimento das demandas judiciais.

Ainda, caso haja descumprimento das medidas exigidas pela lei, a empresa estará sujeita a advertência e multa no valor de até 10% do faturamento do grupo econômico no Brasil no seu último exercício. A arrecadação será destinada ao Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Fundeb) para realização de ações de educação e alfabetização digital.

Seguindo o escopo deste trabalho, o PL possui importantes artigos a serem observados, quais sejam: o 12 e 30. Estes versam respectivamente sobre moderação de conteúdo e sobre o instituto da autorregulação regulada. Nestes dois pontos vê-se bastante a influência da NetzDG alemã, o que é pode ser considerado um fator positivo e importante pela

possibilidade de amadurecimento e melhoria do PL com a verificação da execução prática e lições aprendidas realizada por aquele Estado.

O art. 12 que apresenta os requisitos para a moderação de conteúdo, e estabelece que deve ser garantido acesso à informação e à liberdade de expressão, indo ao encontro do art. 5º incisos IV, XIV da CF (BRASIL, 1988), para que os usuários da plataforma tenham conhecimento adequado sobre condutas permitidas e proibidas e em caso de moderação deve ser explanado como se dará o processo e os meios adequados para recurso caso o usuário não concorde com o resultado. As medidas de moderação devem ser informadas aos usuários de forma a terem a possibilidade de exercer seu direito de defesa respeitados os prazos legais estabelecidos.

O texto ainda prevê situações em que os provedores estão dispensados de notificarem aos usuários se verificarem riscos de dano imediato de difícil reparação ou de violação a direitos de crianças e adolescentes, entre outras situações previstas no PL. Ainda, caso a moderação seja considerada equivocada, deve haver reparação por parte do provedor, no âmbito e nos limites técnicos do serviço. Deverá garantir, ainda, ao ofendido o direito de resposta na mesma medida e alcance do conteúdo considerado inadequado, percebe-se aqui um alinhamento ao art. 5º inciso V da Constituição Federal (BRASIL, 1988). O direito à liberdade de expressão é enfatizado com estas abordagens, permitindo ao usuário transparência e meios para pleitear a solução de discordância com o resultado.

Contudo, em diagnóstico realizado por Maranhão et al. (2021) sobre a moderação de conteúdo no PL2630/2020 há algumas lacunas que precisam ser solucionadas, dessa forma apresenta a seguinte análise:

(...)Faltam regras de transparência e prestação de contas naquilo que ocorre no processo de tomada de decisão pela plataforma, seja em relação aos moderadores, seja em relação aos algoritmos empregados, bem no que diz respeito à disponibilização de mecanismos claros para reclamação para os usuários ofendidos e ao empenho dos provedores no sentido de mitigar impactos do conteúdo nocivo que foi propagado na rede(...) (MARANHÃO et al., 2021)

Já a autorregulação regulada, nos moldes do que foi apresentado na exposição da NetzDG, é trabalhada no art. 30 do projeto de lei, esse o modelo se harmoniza com o papel indutor do Estado regulador da atividade econômica, registrado no art. 174 da CF

(MARANHÃO; CAMPOS, 2022). Assim, os provedores podem se organizar e criar instituto de autorregulação regulada para haja transparência na moderação de conteúdo, sejam definidos procedimentos claros e estabelecidos canais para onde serão realizadas as reclamações, além de ouvidoria independente com a finalidade de receber críticas e avaliar as atividades do instituto. Há o dever de independência e capacidade técnica dos recursos humanos que fazem as avaliações para moderação de conteúdo e o instituto poderá aprovará resoluções e súmulas de modo a regular seus procedimentos de análise.

O instituto de autorregulação regulada deve ser certificado pelo Estado, sendo esse procedimento uma das responsabilidades do Conselho de Transparência e Responsabilidade na Internet. O instituto poderá emitir relatórios para o Conselho expondo dados de moderação de conteúdo, bem como informações de monitoramento de volume de conteúdo compartilhado pelos usuários dos serviços de mensageria privada.

No tocante a constitucionalidade da lei, há necessidade de maiores estudos, visto o grande número de emendas ao projeto, dessa forma, este trabalho procura limitar-se à moderação de conteúdo e o instituto da autorregulação regulada. Porém, é importante observar alguns pontos, pois o texto do projeto de lei foi incrementado com questões que vão além do combate à desinformação e, hoje, seu escopo representa uma grande mudança nas leis sobre a Internet no Brasil, de certo cria-se a necessidade de maior debate acerca do tema.

O Instituto de Tecnologia e Sociedade do Rio, ITS RIO (2022) elencou alguns problemas destas emendas, como por exemplo, há uma consagração da imunidade parlamentar nas redes sociais, porém pode causar algumas situações adversas. Sabe-se que o art. 53 da CF (BRASIL, 2020) garante que deputados e senadores são invioláveis, penal e civilmente, por suas opiniões, palavras e votos, porém isso pode acabar promovendo uma espécie de salvo conduto para ampliar a disseminação de desinformação nas redes sociais pelo desencorajamento da realização de moderação de conteúdo para estes políticos, assim poderia haver uma violação do princípio da igualdade *caput* do art. 5º da CF (BRASIL, 2020).

Relata, ainda, entre os exemplos, o acréscimo ao art. 3º da lei com inciso III – “respeito ao usuário em sua livre formação de preferências políticas e de uma visão de mundo pessoal”. A “visão de mundo pessoal” vai de encontro ao próprio projeto, pois nasceu, entre outros objetivos, para combater a desinformação, e essa redação dificultará a responsabilidade jurídica dos que atuam para desvirtuar a informação nas redes sociais.

Seguindo para o escopo deste trabalho, em análise dos professores Maranhão e Campos (2022) há o entendimento que a instituição da autorregulação regulada é adequada aos preceitos constitucionais, pois com a indução aos provedores de redes sociais para criação de

mecanismos para identificar e classificar conteúdos com notícias fraudulentas, e assim privilegiar, nestas redes, o acesso aos conteúdos corretos, promove a mitigação dos efeitos de desinformação decorrentes da falsificação. São embaraços à comunicação de informação, neste caso cancelados pelo Estado, e para o combate a desinformação. Ainda tecem o seguinte comentário:

(...) Não há, por outro lado, a imposição de restrições à informação, vedada constitucionalmente, uma vez que não há previsão de bloqueio de acesso ou retirada de conteúdo interpretado como lesivo, competência que fica resguardada ao Poder Judiciário, o que está em linha, aliás, com a previsão de *safe harbour* no Marco Civil da Internet (Lei 12.965/14, arts. 18 e 19). Atribuir aos provedores de redes sociais o dever de excluir conteúdo interpretado como lesivo, independentemente de determinação judicial, poderia implicar responsabilização desses mesmos provedores por violações à liberdade de expressão, caso o Poder Judiciário traga entendimento diverso sobre sua qualificação como nocivo(...) (MARANHÃO; CAMPOS, 2020, p. 330)

Até a conclusão deste trabalho o Projeto de Lei encontrava-se aguardando constituição de comissão especial pela Mesa Comissão de Ciência e Tecnologia, Comunicação e Informática. Contudo, houve uma série de audiências públicas realizadas pelo GT (Grupo de Trabalho)³¹ coordenado pela deputada Bruna Furlan e a relatado pelo deputado Orlando Silva, onde foi possível trazer diversos especialistas para debaterem o tema.

Em que pese o trabalho já realizado para elaboração desta lei, a grande quantidade de emendas incluídas cria a necessidade de mais debates sobre o tema e maior atenção quanto a constitucionalidade de seus artigos. Essas alterações dificultam a previsão de como será a redação final do PL ainda mais com a aproximação das eleições de 2022 que pode dificultar ainda mais o andamento do processo legislativo para sua conclusão.

Tanto a NetzDG quando o PL2630/2020 são exemplos que como os Estados estão se estruturando para avançar no tema de moderação de conteúdo em plataformas de redes sociais, buscando em tese garantir a liberdade de expressão e ao mesmo tempo produzir barreiras para a estruturação de campanhas de desinformação. Neste sentido, o capítulo teve como objetivo realizar uma exposição no direito comparado, em meio há algumas reflexões, sobre a necessidade de regulação das plataformas de redes sociais e para isso descreveu os

³¹ As atividades e registros das audiências públicas convocadas pelas Grupo de Trabalho podem ser acessadas via portal da Câmara dos Deputados. Disponível em: <https://www2.camara.leg.br/atividade-legislativa/comissoes/grupos-de-trabalho/56a-legislatura/aperfeicoamento-da-legislacao-brasileira-internet>. Acesso em: 30 de maio de 2022.

principais pontos relevantes a este trabalho com experiência alemã e o PL2630/2020 enfatizando a utilização da autorregulação regulada.

A seguir é apresentada a conclusão deste estudo, que se valeu de observações qualitativas, pela verificação da legislação, doutrina, jurisprudência, projetos de lei e documentação técnica, para se alcançar de maneira dedutiva a problemática estabelecida para regulação das plataformas de redes sociais e a preservação da liberdade de expressão afetada pela moderação de conteúdo.

5 CONCLUSÃO

Este trabalho teve como objetivo discutir um tema bastante caro às Democracias modernas que são as consequências da revolução digital e suas novas formas de comunicação com as plataformas de redes sociais. Ao mesmo tempo que são perceptíveis os ganhos de acesso instantâneo à informação, vê-se por outro lado, o quão prejudicial é a desinformação.

Para isto, foram apresentados conceitos importantes sobre a transformação da esfera pública, nos novos meios de comunicação em massa na Internet, que ultrapassam as barreiras geográficas para distribuir a informação de maneira instantânea, e assim explicar o quão importantes são hoje as plataformas de redes sociais em termos de tamanho e influência na sociedade hiperconectada.

Foi expandido como o princípio da liberdade de expressão é tradicionalmente amparado no ordenamento jurídico brasileiro e sua característica preferencial em relação aos outros princípios. Viu-se também alguns conceitos de desordem informacional e como estão estruturadas as etapas para a elaboração de campanhas de desinformação em massa nas redes sociais. Isso leva a necessidade de que sejam desenhadas soluções para que para que o conteúdo exposto nestas redes seja moderado em caso de necessidade, como por exemplo, discursos fraudulentos e ataques à democracia. Contudo, é preciso ter cuidado em se retirar um conteúdo livre de ilegalidade e assim atentar contra a liberdade de expressão.

Ao analisar as soluções legais propostas para moderação de conteúdo especificamente a NetzDG alemã e o PL2630/2020 brasileira foi possível observar que a legislação alemã encontra-se consolidada e trouxe avanços que podem ser utilizados para construção de legislações em outros países, as críticas em relação ao excesso de moderação que poderia ocasionar violações ao direito a liberdade de expressão não foi consumado na prática, pois nestes anos de execução não houve relatos importantes registrados de *overblocking*.

O PL2630/2020 é bastante influenciado pela legislação alemã, porém tem um atingimento mais amplo pois trata também de serviços de mensageria, além das plataformas de redes sociais. Contudo, apesar de já bastante discutida por especialistas em função das audiências públicas convocadas pelo Grupo de Trabalho, houve um número expressivo de emendas desde sua proposição, o que cria a necessidade de maior debate acerca do tema, mais ainda com uma preocupação de sua constitucionalidade. Há uma perceptível dificuldade em se alcançar um consenso político para o PL avançar até as próximas eleições de 2022, isso pode ocasionar em uma diminuição do combate a desinformação.

Ambas as legislações têm uma preocupação com a necessidade de se deixar transparente os procedimentos, motivações e resultados da moderação de conteúdo, além de garantir o direito recorrer em caso de não concordância. Isso torna-se bastante positivo para proteção à liberdade de expressão e sob outro aspecto promovem ferramentas que podem atuar de forma rápida na moderação de conteúdo manifestadamente ilegal como casos de desinformação.

A instituição da autorregulação regulada tem se mostrado adequada à moderação de conteúdo como observado na experiência alemã. Pois permite que uma instituição independente, acreditada pelo Estado, atue nos casos mais complexos aumentando o potencial de acerto na determinação da legalidade ou ilegalidade do conteúdo. Apesar de haver previsão no PL brasileiro, não é possível afirmar que será adequado ao contexto nacional, porém o instituto em si tem se mostrado uma elegante solução para tratar o assunto da moderação.

Assim, baseando-se na experiência alemã, é possível concluir que apesar do cenário de desinformação em massa veiculado nas plataformas de rede social, a liberdade de expressão tem sido preservada na moderação de conteúdo na aplicação da lei, garantido principalmente por meio das obrigações de transparência em todo o processo da solução.

Cabe também concluir que é necessário que se avance na legislação brasileira, principalmente para que haja uma organização padronizada na moderação de conteúdo e se acolham as melhores práticas observadas na legislação alemã, bem como se solucionem as lacunas no projeto de lei indicados pelos especialistas.

Uma dificuldade encontrada na realização deste trabalho é que apesar de bastante relevante, o tema parece ser de difícil unidade política em qualquer Estado, porém no cenário brasileiro há uma certa dificuldade trazida pela polarização política que muito se dá nas plataformas de redes sociais o que acarreta mudanças das mais diversas na redação, às vezes descaracterizando-a, e desinteresse em se alcançar uma celeridade para aprovação do projeto.

Para trabalhos futuros pretende-se exercitar uma maior profundidade no exame tema e também sejam incorporadas análises de outras legislações como, por exemplo, o *Digital Service Act* proposto pela União Europeia e que encontra-se bastante avançada.

REFERÊNCIAS

ABBOUD, Georges; CAMPOS, Ricardo. **A autorregulação regulada como modelo do Direito proceduralizado: Regulação de redes sociais e proceduralização**. In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2020.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2020]. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm. Acesso em: 04 jun. 2022.

BITTMAN, Ladislav. **A KGB e a desinformação Soviética**. 1. ed. Campinas: Vide Editorial, 2019.

BRANCO, Dácio C. **7 em cada 10 brasileiros se informa por redes sociais — e isso afeta a segurança**. Canaltech, 13 de out. de 2021. Disponível em: <https://canaltech.com.br/seguranca/7-em-cada-10-brasileiros-se-informa-por-redes-sociais-e-isso-afeta-a-seguranca-198668/>. Acesso em: 27 de nov. 2021.

BRASIL. Marco Civil da Internet. Lei 12.964/2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em: 18 de maio. 2022

BRANDWATCH, **Brandwatch's 50 Most Influential People on Twitter 2021**, 2021. Disponível em: <https://www.brandwatch.com/blog/50-most-influential-people-twitter/>. Acesso em: 27 de nov. 2021.

CAMPOS, Ricardo. **Transformação da esfera pública motivou nova lei alemã de Internet**. In: Consultor Jurídico, 4 de jul. de 2017. Disponível em: <https://www.conjur.com.br/2017-jul-04/opiniao-transformacao-social-motivou-lei-alema-internet>. Acesso em: 21 de nov. de 2021.

CAMPOS, Ricardo; GRINGS, Maria Gabriela. **Sandbox e autorregulação regulada: uma análise comparada Europa-Brasil** In: Legal Innovation - O Direito do Futuro e o Futuro do Direito. Coordenadores: Rony Vainzof; Danielle Serafino e Aline Steinwascher. São Paulo: Thomson Reuters Brasil, 2022.

CLÈVE, Clèmerson Melin. **Direito Constitucional Brasileiro: TEORIA DA CONSTITUIÇÃO E DIREITOS FUNDAMENTAIS - Vol 1** São Paulo: Revista dos Tribunais, 2022

CUEVA, Ricardo Villas Boas. **Alternativas para a remoção de fake news das redes sociais**. In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2020.

DATAREPORTAL, **Digital Around the World**, 2021a. Disponível em: <https://datareportal.com/global-digital-overview>. Acesso em: 27 de nov. 2021

_____. **Digital 2021: Brazil**, 2021b. Disponível em: <https://datareportal.com/reports/digital-2021-brazil>. Acesso em: 27 de nov. 2021

EIFERT, Martin. **A Lei Alemã para a Melhoria da Aplicação da Lei nas Redes Sociais (NetzDG) e a Regulação da Plataforma**. In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2020.

EIFERT, Martin. **Prefácio da 3º ed.** In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2022.

ESTADÃO CONTEÚDO. **Facebook atinge pela primeira vez US\$ 1 trilhão em valor de mercado nos EUA**, 29 de jun de 2021. Disponível em: <https://www.infomoney.com.br/mercados/facebook-atinge-pela-primeira-vez-us-1-trilhao-em-valor-de-mercado-nos-eua/>. Acesso em: 27 de nov. 2021

FONTENELES, Samuel Sales. **Direitos Fundamentais**. Salvador: Editora JusPodivm, 2021.

GOOGLE **Transparency Report. Cumprimento das diretrizes da comunidade do YouTube**. Disponível em: <https://transparencyreport.google.com/youtube-policy/removals>. Acesso em: 22 mai.2022.

GOOGLE **Transparency Report. Solicitações governamentais de remoção de conteúdo**. Disponível em: <https://transparencyreport.google.com/government-removals/overview>. Acesso em: 22 mai.2022.

GRIMMELMANN, James. **The Virtues of Moderation**. Yale J.L. & Tech, 2015. Disponível em: <https://scholarship.law.cornell.edu/facpub/1486/>

HABERMAS, Jürgen. **Mudança Estrutural da Esfera Pública: Investigações sobre uma categoria da sociedade burguesa**. São Paulo: Unesp, 2014.

_____. **Direito e democracia: entre facticidade e validade**. Volume II. Rio de Janeiro: Tempo Brasileiro, 1997.

FARIA, José E. **A democracia e as redes sociais**. Estadão, São Paulo, 23 de ago de 2019. Disponível em: <https://estadodaarte.estadao.com.br/a-democracia-e-as-redes-sociais/> Acesso em: 22 de nov. de 2021.

ITS RIO. **10 pontos de atenção sobre o PL das Fake News (PL 2630/20)**. In: Instituto de Tecnologia e Sociedade do Rio. ITS FEED. Rio de Janeiro, 29 mar. 2022. Disponível em: <https://feed.itsrio.org/10-pontos-de-atencao-sobre-o-pl-das-fake-news-pl-2630-20-a1420e502c66>. Acesso em: 5 jun. 2022.

KLONICK, Kate. **The New Governors: The People, Rules, and Processes Governing Online Speech**. Harvard Law Review, [s. l.], v. 131, p. 1.598-1.670, 2018

MARANHÃO, Juliano; CAMPOS, Ricardo. **Fake News e autorregulação regulada das redes sociais no Brasil: fundamentos constitucionais**. In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2022.

MARANHÃO, Juliano; CAMPOS, Ricardo; KETTEMANN, Matthias; ABRUSIO, Juliana; SARTOR, Giovanni. **Como regular a moderação privada de conteúdo nos novos espaços públicos?** In: Consultor Jurídico, 1 de set. de 2020. Disponível em: <https://www.conjur.com.br/2020-set-01/direito-digitalcomo-regular-moderacao-privada-conteudo-novos-espacos-publicos>. Acesso em: 21 de maio. de 2022.

MARANHÃO, Juliano et al. **REGULAÇÃO DE “FAKE NEWS” NO BRASIL**. In: Instituto Legal Grounds. REGULAÇÃO DE “FAKE NEWS” NO BRASIL. [S.l.]. 6 out. 2021. Disponível em: <https://institutolgpd.com/blog/regulacao-de-fake-news-no-brasil/>. Acesso em: 4 jun. 2022.

MENDES, G. F.; SARLET, I. W.; LTDA, I. C. E. P.; LTDA, I. C. E. P.; CANOTILHO, J. J. G.; LEONCY, L. F.; STRECK, L. L. **Comentários à Constituição do Brasil**. 2. ed. São Paulo: Saraiva, 2017. E-book.

NAPOLI, Philip M. **Social media and the public interest : media regulation in the disinformation age**. New York : Columbia University Press, 2019.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. **Declaração Universal dos Direitos Humanos**, 1948. Disponível em: <https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos>. Acesso em: 8 Maio. 2022.

RECUERO, Raquel. **Redes sociais na internet**. Porto Alegre: Sulina, 2009

SARMENTO, Daniel. Art. 5º, IV. In: CANOTILHO, José Joaquim Gomes et al. (orgs.). **Comentários à Constituição do Brasil**. 2. ed. São Paulo: Saraiva, 2018.

SCHNEIER, Bruce. **Toward an Information Operations Kill Chain**, 2019. Disponível em: <https://www.lawfareblog.com/toward-information-operations-kill-chain>. Acesso em: 20 de maio. 2022.

SILVA, José Afonso. **Curso de Direito Constitucional Positivo**. São Paulo: Malheiros, 2011.

STATISTA. **Most popular social networks worldwide as of October 2021, ranked by number of active users**, 2021. Disponível em: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>. Acesso em: 27 de nov. 2021.

STROPPA, Tatiana; CARVALHO, Leticia Redis; GRINGS, Maria Gabriela; ROXO, Tatiana Bhering; OLIVEIRA, Samuel. **A seção 230 do CDA e o artigo 19 do Marco Civil da Internet In: Consultor Jurídico**, 4 de maio. de 2022. Disponível em: <https://www.conjur.com.br/2022-mai-04/direito-digital-secao-230-cda-artigo-19-marco-civil-internet>. Acesso em: 25 de maio. de 2022.

TWITTER. **Transparency. Rules Enforcement.** Disponível em: <https://transparency.twitter.com/en/twitter-rules-enforcement.html>. Acesso em: 21 maio. 2022.

VESTING, Thomas. **A mudança da esfera pública pela inteligência artificial.** In: Fake news e Regulação. Coordenadores: Georges Abboud; Nelson Nery Jr. e Ricardo Campos. São Paulo: Thomson Reuters Brasil, 2020.

WARDLE, Claire.; DERAKHSHAN, Hussein. **INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making.** Council of Europe. [S.l.]. 2017.

WIKIPEDIA. Section 230, 2022. Disponível em: https://en.wikipedia.org/wiki/Section_230. Acesso em: 25 de maio. 2022.

WORLDOMETER. **Current World Population**, 2021. Disponível em: <https://www.worldometers.info/world-population/>. Acesso em: 27 de nov. 2021