

# **V ENCONTRO VIRTUAL DO CONPEDI**

## **DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS II**

**JOSÉ RENATO GAZIERO CELLA**

**AIRES JOSE ROVER**

**FERNANDO GALINDO AYUDA**

Todos os direitos reservados e protegidos. Nenhuma parte deste anal poderá ser reproduzida ou transmitida sejam quais forem os meios empregados sem prévia autorização dos editores.

**Diretoria - CONPEDI**

**Presidente** - Prof. Dr. Orides Mezzaroba - UFSC - Santa Catarina

**Diretora Executiva** - Profa. Dra. Samyra Haydêe Dal Farra Napolini - UNIVEM/FMU - São Paulo

**Vice-presidente Norte** - Prof. Dr. Jean Carlos Dias - Cesupa - Pará

**Vice-presidente Centro-Oeste** - Prof. Dr. José Querino Tavares Neto - UFG - Goiás

**Vice-presidente Sul** - Prof. Dr. Leonel Severo Rocha - Unisinos - Rio Grande do Sul

**Vice-presidente Sudeste** - Profa. Dra. Rosângela Lunardelli Cavallazzi - UFRJ/PUCRio - Rio de Janeiro

**Vice-presidente Nordeste** - Profa. Dra. Gina Vidal Marcilio Pompeu - UNIFOR - Ceará

**Representante Discente:** Prof. Dra. Sinara Lacerda Andrade - UNIMAR/FEPODI - São Paulo

**Conselho Fiscal:**

Prof. Dr. Caio Augusto Souza Lara - ESDHC - Minas Gerais

Prof. Dr. João Marcelo de Lima Assafim - UCAM - Rio de Janeiro

Prof. Dr. José Filomeno de Moraes Filho - Ceará

Prof. Dr. Lucas Gonçalves da Silva - UFS - Sergipe

Prof. Dr. Valter Moura do Carmo - UNIMAR - São Paulo

**Secretarias**

**Relações Institucionais:**

Prof. Dra. Daniela Marques De Moraes - UNB - Distrito Federal

Prof. Dr. Horácio Wanderlei Rodrigues - UNIVEM - São Paulo

Prof. Dr. Yuri Nathan da Costa Lannes - Mackenzie - São Paulo

**Comunicação:**

Prof. Dr. Liton Lanes Pilau Sobrinho - UPF/Univali - Rio Grande do Sul

Profa. Dra. Maria Creusa De Araújo Borges - UFPB - Paraíba

Prof. Dr. Matheus Felipe de Castro - UNOESC - Santa Catarina

**Relações Internacionais para o Continente Americano:**

Prof. Dr. Heron José de Santana Gordilho - UFBA - Bahia

Prof. Dr. Jerônimo Siqueira Tybusch - UFSM - Rio Grande do Sul

Prof. Dr. Paulo Roberto Barbosa Ramos - UFMA - Maranhão

**Relações Internacionais para os demais Continentes:**

Prof. Dr. José Barroso Filho - ENAJUM

Prof. Dr. Rubens Beçak - USP - São Paulo

Profa. Dra. Viviane Coêlho de Séllos Knoerr - Unicuritiba - Paraná

**Eventos:**

Prof. Dr. Antônio Carlos Diniz Murta - Fumec - Minas Gerais

Profa. Dra. Cinthia Obladen de Almendra Freitas - PUC - Paraná

Profa. Dra. Livia Gaigner Bosio Campello - UFMS - Mato Grosso do Sul

**Membro Nato** - Presidência anterior Prof. Dr. Raymundo Juliano Feitosa - UMICAP - Pernambuco

D597

Direito, governança e novas tecnologias II [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Aires Jose Rover; Fernando Galindo Ayuda; José Renato Gaziero Cella – Florianópolis: CONPEDI, 2022.

Inclui bibliografia

ISBN: 978-65-5648-481-5

Modo de acesso: [www.conpedi.org.br](http://www.conpedi.org.br) em publicações

Tema: Inovação, Direito e Sustentabilidade

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Nacionais. 2. Direito. 3. Governança. V Encontro Virtual do CONPEDI (1: 2022 : Florianópolis, Brasil).

CDU: 34



## **V ENCONTRO VIRTUAL DO CONPEDI**

### **DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS II**

---

#### **Apresentação**

No V Encontro Virtual do CONPEDI, realizado de 14 a 18 de junho de 2022, o grupo de trabalho “Direito, Governança e Novas Tecnologias II”, que teve lugar na tarde de 15 de junho de 2022, destacou-se no evento não apenas pela qualidade dos trabalhos apresentados, mas pelos autores dos artigos, que são professores pesquisadores acompanhados de seus alunos pós-graduandos. Foram apresentados 22 artigos objeto de um intenso debate presidido pelos coordenadores e acompanhado pela participação instigante do público presente na sala virtual.

Esse fato demonstra a inquietude que os temas debatidos despertam na seara jurídica. Cientes desse fato, os programas de pós-graduação em direito empreendem um diálogo que suscita a interdisciplinaridade na pesquisa e se propõe a enfrentar os desafios que as novas tecnologias impõem ao direito. Para apresentar e discutir os trabalhos produzidos, os coordenadores do grupo de trabalho dividiram os artigos em três blocos, quais sejam: a) inteligência artificial; b) proteção de dados pessoais; c) novas tecnologias, internet e redes sociais. Segue os temas principais de cada bloco:

O bloco de trabalhos da inteligência artificial, os artigos levantaram temas como A INTELIGÊNCIA ARTIFICIAL COMO INSTRUMENTO DE EFICIÊNCIA NA ANÁLISE ECONÔMICA DO DIREITO, NO PROCESSO DO TRABALHO, NO REGISTRO DE IMÓVEIS, NO ACESSO À JUSTIÇA. O FUTURO DA INTELIGÊNCIA ARTIFICIAL E SUA REGULAÇÃO. FINALMENTE, A DISCRIMINAÇÃO ALGORÍTIMICA.

O segundo bloco sobre proteção de dados pessoais trouxe temas como ESTUDO COMPARADO SOBRE PROTEÇÃO DE DADOS PESSOAIS, A PROTEÇÃO DOS DADOS DAS CRIANÇAS E DOS ADOLESCENTES, A PUBLICIDADE REGISTRAL DO DIREITO DE PROPRIEDADE, O PAPEL DA AGÊNCIA BRASILEIRA DE INTELIGÊNCIA (ABIN), E A LGPD COMO INDUTORA PARA A TRANSPARÊNCIA NO LEGISLATIVO.

O terceiro bloco, das novas tecnologias, internet e redes sociais congregaram temas como AUTORREGULAÇÃO E O FACEBOOK, A TRANSNACIONALIDADE E O REGISTRO IMOBILIÁRIO, DIMENSÃO JURÍDICA DO OLIMPISMO E AS NOVAS TECNOLOGIAS DE MÍDIA, A RESPONSABILIDADE CIVIL NO MARCO CIVIL DA

INTERNET, INOVAÇÕES TECNOLÓGICAS NO ACESSO À JUSTIÇA, MEIO AMBIENTE E A GOVERNANÇA DIGITAL, MODERAÇÃO DE CONTEÚDO PELAS MÍDIAS SOCIAIS, MOVIMENTOS SOCIAIS DIGITAIS E A DESOBEDIÊNCIA CIVIL E O NET-ATIVISMO.

Os artigos que ora são apresentados ao público têm a finalidade de fomentar a pesquisa e fortalecer o diálogo interdisciplinar em torno do tema “Direito, Governança e Novas Tecnologias”. Trazem consigo, ainda, a expectativa de contribuir para os avanços do estudo desse tema no âmbito da pós-graduação em direito brasileira, apresentando respostas para uma realidade que se mostra em constante transformação.

Os Coordenadores

Prof. Dr. Aires José Rover

Prof. Dr. Fernando Galindo

Prof. Dr. José Renato Gaziero Cella

# MODERAÇÃO DE CONTEÚDO PELAS MÍDIAS SOCIAIS NO COMBATE À DESINFORMAÇÃO

## CONTENT MODERATION SOCIAL MEDIA IN FIGHTING DISINFORMATION

**Gustavo Silveira Borges  
Fabio Jeremias De Souza**

### **Resumo**

O artigo tem por objetivo geral estudar a moderação de conteúdo pelas mídias sociais para verificar algumas das principais iniciativas de combate à desinformação. Apresenta uma análise conceitual e histórica do tema, passando pelas cartas de princípios. Após, buscou-se apresentar como funciona a moderação de conteúdo de desinformação pelas das mídias sociais. Optou-se pela pesquisa qualitativa e pelo método dedutivo. Conclui-se que os direitos humanos são o elo de ligação entre as mídias sociais e o direito de cada um dos milhares de usuários dos países onde as plataformas operam.

**Palavras-chave:** Moderação de conteúdo, Mídias sociais, Desinformação, Direitos humanos, Cartas de princípios

### **Abstract/Resumen/Résumé**

The article has as the general objective the study of the moderation of the content by social medias to verify some of the principal initiatives in the combat of disinformation. It Presents a conceptual and historical analysis of the theme, going through the letters of principles. After, it looked to present how the moderation of disinformation content by social medias work. It opted for the qualitative research and deductive method. In conclusion the human rights are the link between the social medias and the rights of each one of the millions of users on the countries where the platforms operate.

**Keywords/Palabras-claves/Mots-clés:** Content moderation, Social media, Desinformation, Human rights, Letters of principles

## 1. INTRODUÇÃO

A propagação de mentiras remonta tempos imemoriais. No entanto, o avanço da *internet*, das novas tecnologias e do crescente número de usuários, fez com que o problema avançasse numa escala assustadora, a ponto de comprometer a própria democracia. Esse é o maior desafio regulatório dos novos tempos (JORGENSEN, 2021, p. 1).

As mídias sociais e suas plataformas estão amparadas nos princípios da livre iniciativa e concorrência. Contudo, a crescente adesão da população gera repercussões na esfera coletiva. Assim, tratar do tema implica reconhecer o caráter privado das mídias sociais, sem descuidar do impacto social da sua atuação. Ainda, deve ser ressaltado que o ambiente está cada vez mais concentrado num pequeno número de mídias sociais, sendo que no início, plataformas como Facebook, Youtube e Twitter eram tidas como “tecnologias de libertação”, com o escopo de permitir a liberdade de expressão, mobilizar protestos e fiscalizar governos, ao mesmo tempo em que surgiram as inúmeras controvérsias públicas (SANDER, 2021, p. 160).

O avanço exponencial da tecnologia da informação e comunicação (TIC), trouxe muitas oportunidades de compartilhamento e expressão de opiniões. Nesse contexto, deve ser considerado que as consequências das implicações da responsabilidade corporativa devem respeitar os direitos humanos universais, sendo essa a base do relacionamento entre as mídias sociais e seus usuários (BUHMANN; OLIVERA, 2020, p. 5).

Nesse sentido, o presente trabalho tem por objetivo geral, verificar, a partir do estudo do histórico de moderação de conteúdo pelas mídias sociais, assim como dos procedimentos e instrumentos seguidos, quais seriam algumas das principais iniciativas de combate à desinformação por meio da moderação de conteúdo.

No primeiro capítulo, será apresentado um itinerário histórico acerca da moderação de conteúdo pelas mídias sociais, passando pelas cartas de princípios, com especial ênfase na “Chamada de Paris”, que enfatiza a questão da desinformação. Ainda no primeiro capítulo, serão apresentadas algumas iniciativas governamentais, como a NetzDG (*Network Enforcement Act*) alemã, O Relatório Francês (*Creating a French framework to make social media platforms more accountable*) e o *Online Harms White Paper*, do Reino Unido.

O segundo capítulo trata da moderação de conteúdo pelas mídias sociais, com ênfase na desinformação. Além de abordar os *standards* de direitos humanos, buscou-se apresentar os padrões da comunidade do Facebook no combate à desinformação, maior mídia social do mundo, com o objetivo de apresentar os parâmetros hoje utilizados.

Deste modo, o problema da presente pesquisa é o seguinte: a partir do estudo do histórico de moderação de conteúdo pelas mídias sociais, assim como dos procedimentos e instrumentos seguidos, quais seriam algumas das principais iniciativas de combate à desinformação por meio da moderação de conteúdo?

Este artigo se justifica pela atualidade do tema, pelo avanço das novas tecnologias e pela expansão das mídias sociais, cada vez mais constituídas como meio de obtenção e difusão da informação. Deve-se considerar ainda que as mídias sociais mediam as relações de poder e controle entre diferentes atores, com foco em dois pontos, a moderação de conteúdo e a proteção de danos. (SANDER, 2021, p. 161).

Escolheu-se a análise da moderação de conteúdo, especialmente no que se refere à desinformação, porém, deve ser ressaltado que um assunto (moderação) está ligado no outro (proteção de dados), na medida que foi justamente através do garimpo e tratamento de dados, que o mundo acompanhou uma grande experiência de manipulação do eleitor: o caso *Cambridge Analytica*. Estudar o tema tem como intuito reunir elementos para permitir o aperfeiçoamento da defesa da própria democracia, a partir do momento em que se constatou que a difusão articulada da desinformação pode significar a manipulação da vontade do eleitor.

A presente pesquisa adotou o método de abordagem dedutivo, eis que partiu de premissas gerais a fim de chegar a uma conclusão particular. Por sua vez, a técnica de pesquisa a bibliográfica, uma vez que se utilizou de livros e artigos científicos já elaborados.

## **2. A DESINFORMAÇÃO E O ITINERÁRIO HISTÓRICO DA MODERAÇÃO DE CONTEÚDO DAS MÍDIAS SOCIAIS**

A desinformação não é algo novo e, conforme observa Barbosa (2021, p. 758), a maioria dos estudos não considera dessa forma, nem considerada que não são exclusividade do mundo digital, embora reconheça-se que a massificação é fruto da democratização da *internet*.

Como a mentira, a busca pela verdade também não é algo novo. Platão sugeriu que existem verdades absolutas que se escondem do homem, mas que existem. Aguardam a sua descoberta por meio do conhecimento, da iluminação, passagem retratada na alegoria da caverna que, em apertada síntese, ensina que a verdade absoluta não será revelada pelos sentidos, mas sim pela razão (NEISSER, 2016, p. 237).

Apesar de não se estar lidando com algo novo, é inegável o aumento exponencial do problema, a partir da democratização da internet. Como experiência recente, deve ser considerado o fenômeno ocorrido com o *Brexit*, em junho de 2016. Convocado pelo ex-primeiro

ministro David Cameron, o *Brexit* foi uma consulta popular que renegociava a permanência do Reino Unido na União Europeia (UE). Os grupos favoráveis à saída da UE elegeram como inimigo o impacto negativo da imigração, sobretudo no que se refere aos refugiados sírios. O grupo que pleiteava a permanência, por outro lado, contou com a participação na campanha de figuras importantes, como os artistas, através de um discurso pautado nos avanços obtidos com livre trânsito, o comércio e moeda única. O resultado do referendo, pela saída do Reino Unido da EU, surpreendeu. Anos de construção de uma Europa unificada, foram descartados por um discurso raso e superficial. No episódio, foi constatada a participação de robôs, a observação de sentimentos e novos métodos de previsão de resultados eleitorais (hashtags, palavras-chave e menções nas redes). As abordagens sugeriram uma relação entre as redes sociais, sobretudo o *Twitter* com o resultado do referendo (REDES SOCIAIS NAS ELEIÇÕES 2018, FGV, p. 11).

Da mesma forma, a eleição americana de 2016 é exemplo interessante para a análise da influência da desinformação, sobretudo em processo eleitorais. Donald Trump venceu as eleições americanas após uma campanha ostensivamente tecnológica, permeada de dúvidas quanto à sua integridade. A campanha contou com a contratação da empresa *Cambridge Analytica*:

“A Cambridge Analytica, multinacional britânica, utilizou como estratégia de atuação uma metodologia que busca traçar a personalidade dos indivíduos com base em preceitos clássicos de psicologia e psicometria, aliando essas técnicas aos rastros digitais que são deixados cotidianamente pelas pessoas em suas atividades comuns, “como perfis em redes sociais, GPS de locais visitados, dados de uso dos serviços públicos e compras on line” (RAIS, 2018, p. 77).

A empresa utilizou duas frentes para o seu intento, através da coleta de dados e utilizando o *big data*: uma *off line*, com envio de materiais para residências e programas de TV e outra patrocinando publicidade direcionada nas redes sociais, invariavelmente distorcendo a verdade.

Embora as notícias falsas tenham prevalecido durante vários anos antes das eleições presidenciais americanas de 2016, o tema só recentemente se tornou interessante para pesquisas de estudiosos, pois antes da eleição elas não haviam sido muito persuasivas (CHIRWA, 2021, p. 60). A inovação trazida pelo caso *Cambridge Analytica* e que fez toda a diferença é que o objetivo principal dos estudos era a análise comportamental dos indivíduos, o que só é possível com o avanço da tecnologia (FORNAISER; BECK, 2019, p. 186).

Há uma outra questão que revela que as notícias falsas possuem um engajamento maior. A checagem dos fatos é realizada, infelizmente, por uma minoria, resultando no

problema mais devastador do tempo da pós-verdade: a desvalorização e deslegitimação de vozes especializadas e instituições (AIETA, 2020, p. 232).

Enfim, a análise do tema requer um duplo cuidado. O primeiro, de combater os efeitos nocivos e exponenciais da desinformação. De outro lado, equacionar e implementar no combate, regras claras de modo impedir o cerceamento à liberdade de expressão. Antes do avanço das mídias sociais, a maior ameaça à liberdade de expressão era justamente o poder do Estado, de usar a punição como medida capaz de regular a difusão de ideias contrárias, questionadoras e fiscalizadoras. (SANDER, 2021, p. 166)

Já no que se refere a desinformação, importante a distinção sobre as categorias de notícias falsas apresentadas pelo *Policy Department for Citizen Rights Constitutional Affairs*: a) *Missinformation*: Informação falsa, sem objetivo de causar lesão; b) *Disinformation*: Divulgação de informação falsa de forma deliberada para causar lesão; c) *Malinformation*: divulgação genuína para causar dano (BARBOSA, 2021, p. 759).

A desinformação gera receita, porque histórias negativas geram mais acessos. De acordo com Carvalho (2022, p. 177), para atingir um número expressivo de audiência, princípios éticos do jornalismo tradicional e a qualidade do conteúdo são desprezados.

O somatório desses fatores faz com que o tema do combate à desinformação ganhe contornos relevantes, mormente diante da constatação de que as instituições e a convivência democrática podem ser abaladas a partir de uma lógica em que a verdade e a razão não importam, o que importam são os índices de audiência (D'ANCONA, 2018, p. 23). Ainda, de acordo com Empoli (2019, p. 9) “*para os adeptos dos populistas, a verdade dos fatos, um a um, não conta. O que é verdadeiro é a mensagem no seu conjunto, que corresponde aos seus sentimentos e suas sensações*”.

Assim, reunindo meios cada vez mais sofisticados, a utilização da desinformação como instrumento político e econômico, vem ganhando corpo com contornos cada vez mais sofisticados, de forma que, invariavelmente, vem à tona a discussão a respeito da responsabilidade das redes diante deste cenário, cujo histórico será visto a seguir.

## **2.1. Contornos históricos da moderação de conteúdo pelas mídias sociais**

As principais mídias sociais hoje difundidas tiveram origem nos anos 2000 e possuem sede nos Estados Unidos. Porém, já nos anos 90 tivemos os primeiros casos judiciais envolvendo a discussão a respeito da responsabilidade dos intermediários, com destaque para dois casos emblemático. No ano de 1991 o Tribunal Distrital dos EUA para o Distrito Sul de

Nova York decidiu o caso *Cubby v CompuServe* (JUSTIA US LAW, 1991). Nesse caso, a plataforma foi acionada pela empresa Cubby. A Ré *CompuServe* desenvolvia e fornecia produtos e serviços relacionados a computadores, incluindo o *CompuServe Information Service* ("CIS"), espécie de "biblioteca eletrônica" que os assinantes podem acessar de um computador pessoal ou de um terminal.

O fundamento fático da ação residia no ataque a um produto da *Cubby*, chamado *Skuttlebut* (um banco de dados para publicar e distribuir eletronicamente notícias e fofocas nos noticiários de televisão e indústrias de rádio), atacado por uma concorrente através de declarações falsas e difamatórias, através de um fórum da *CompuServe*. O Tribunal Distrital entendeu que a empresa *CompuServe* não possuía relação direta com o conteúdo, tanto por não possuir relação comercial com os terceiros que publicavam conteúdo, como por não promover controle prévio sobre a veiculação. (ESTARQUE; ACHEGAS; BOTTINO; PERRONE, 2021, p. 17).

O segundo foi o caso *Stratton Oakmont v Prodigy Services*, julgado em 1995, pela Suprema Corte de Nova York. A empresa *Stratton Oakmont, Inc.*, do ramo de investimentos em valores mobiliários teria praticado atos fraudulentos na oferta pública de ações de uma determinada empresa. A *Prodigy Services* seria uma desenvolvedora da plataforma digital em que foram divulgados os atos fraudulentos. O julgado, inclusive, rememora o caso *Cubby x CompuServe*:

*“A principal distinção entre CompuServe e PRODIGY é dupla. Em primeiro lugar, a PRODIGY apresentou-se ao público e aos seus membros como controlando o conteúdo dos seus boletins informáticos. Em segundo lugar, a PRODIGY implementou esse controle por meio de seu programa de triagem automática de software e das Diretrizes que os Líderes do Conselho devem aplicar. Ao utilizar ativamente tecnologia e mão de obra para excluir notas de seus quadros de avisos de computador com base em ofensividade e "mau gosto", por exemplo, a PRODIGY está claramente tomando decisões quanto ao conteúdo (ver, Miami Herald Publishing Co. v. Tornillo, supra), e tais decisões constituem controle editorial. ( ID.) ”( H2O, 1995)*

Dos casos acima relatados, temos que através da utilização da jurisprudência do *common law* em relação às plataformas de mídias sociais, foi estabelecida a distinção entre um editor (que opta por moderar conteúdo) de um mero distribuidor. No primeiro caso, a plataforma passa a ser responsabilizada pelo conteúdo. Assim, controlar o conteúdo poderia gerar a responsabilidade civil, ao passo que as mídias sociais foram naturalmente se apresentando como simples distribuidoras, prática que trouxe uma série de consequências (ESTARQUE; ACHEGAS; BOTTINO; PERRONE, 2021, p. 18)

Com o objetivo de harmonizar esses dois precedentes, foi editada a Seção 230, da *Communications Decency Act*, de 1996. Pela redação do artigo *“nenhum provedor ou usuário*

*de serviço interativo de computador será tratado como editor de qualquer informação fornecida por terceiros*". (ESTARQUE, ACHEGAS, BOTTINO; PERRONE; 2021; p. 18)

Partiu-se do pressuposto de que as plataformas não devem ser caracterizadas como editora, além de conter o dispositivo contém uma cláusula chamada de “cláusula do bom samaritano”, que estatui a extensão da imunidade às mídias que optem por moderar conteúdo.

Importante considerar ainda, nesse contexto, que a liberdade de expressão nos Estados Unidos possui um valor superlativo, inspirada na Primeira Emenda à Constituição. Em síntese, a Primeira Emenda estipula que o Congresso não poderá limitar a liberdade de expressão, assim com a liberdade religiosa, liberdade de reunião e direito de petição aos órgãos públicos.<sup>1</sup>

Porém, mesmo tendo as empresas sede nos Estados Unidos, por pressão dos usuários e por questão de sobrevivência no mercado, as plataformas optaram por moderar o conteúdo e garantir a segurança do usuário, ainda que de forma tímida. O mercado exigiu, em contraponto à liberdade ilimitada de expressão, que o espaço não fosse transformado numa “terra de ninguém”, onde os usuários não se sentissem acolhidos e seguros. Contudo, considerando os conflitos que foram surgindo com a proliferação e democratização das redes sociais, tanto as entidades da sociedade civil e, sobretudo, as entidades governamentais, não se conformaram com a autorregulação, passando a exigir políticas ainda mais rígidas das plataformas. Surgiram diversas iniciativas, como as cartas de princípios.

## **2.2. Cartas de Princípios**

De importância para a análise do tema, não se pode deixar de mencionar as cartas de princípios, que são iniciativas da sociedade civil organizada, por vezes com participação governamental, com o intuito de implementar diretrizes para a moderação de conteúdo, com foco na proteção do usuário, nos *standards* de direitos humanos e na liberdade de expressão. (ESTARQUE; ACHEGAS, BOTTINO; PERRONE; 2021; p. 29)

A Manila Principles, de 2015, tem como objetivo propor medidas que garantam o direito à liberdade de expressão, a partir de recomendações baseadas em instrumentos de direitos humanos internacionais, assim como em outros marcos internacionais legais. (MANILA PRINCIPLES, 2015). Em resumo, a carta considera que os intermediários devem ser protegidos

---

<sup>1</sup> “O Congresso não fará nenhuma lei a respeito do estabelecimento de uma religião, ou proibindo seu livre exercício; ou cerceando a liberdade de expressão ou de imprensa; ou o direito do povo de se reunir pacificamente e de solicitar ao governo a reparação de queixas”. Primeira Emenda. Disponível em [https://www.law.cornell.edu/constitution/first\\_amendment](https://www.law.cornell.edu/constitution/first_amendment) Acesso em: 01 mar. 2022.

por lei da responsabilização, com regras claras, objetivas, excluindo a reponsabilidade quando não tenham realizado qualquer modificação no conteúdo, sendo que a remoção deve vir acompanhada de uma ordem judicial. A carta prestigia também o devido processo legal, assim como estabelece que as leis, ordens e as práticas de restrição de conteúdo, devem levar em conta a necessidade e proporcionalidade. Por fim, ela menciona os deveres de transparência e de prestação de contas.

Aliás, uma primeira crítica quanto aos desafios trazidos pelas novas tecnologias em defesa dos direitos humanos é a falta de transparência, considerando que o desenvolvimento das mídias tem sido "uma caixa preta", porque os usuários desconhecem os processos de tomada de decisão. (IZA, 2021, p. 3)

Já os princípios de Santa Clara (*Santa Clara Principles*), tiveram origem em 2018, na *Content Moderation at Scale* nos Estados Unidos. Na oportunidade, um grupo de estudiosos em direitos humanos e especialistas desenvolveu um conjunto de princípios sobre a melhor forma de obter transparência e responsabilidade quanto à moderação de conteúdo nas redes sociais. Doze grandes empresas aderiram aos Princípios de Santa Clara, com destaque para a *Apple, Facebook (Meta), Google, Reddit, Twitter e Github* (SANTA CLARA PRINCIPLES, 2018).

Inicialmente, existiam os Princípios de Santa Clara 1.0. e, como o passar dos tempos e a necessidade de aprofundamento, surgiram os Princípios de Santa Clara 2.0. Quanto ao Princípios de Santa Clara 1.0., preveem a necessidade de veiculação do número total de postagens removidas e de contas suspensas, o dever de avisar cada usuário cujo conteúdo for removido ou suspenso, e ainda que as empresas devem fornecer oportunidades para recursos (SANTA CLARA PRINCIPLES, 2018).

Com relação aos Princípios de Santa Clara 2.0., que remonta os anos de 2020 e 2021, foi dividida em princípios fundamentais e operacionais. Os princípios fundamentais dividem-se em cinco: a) Direitos Humanos e devido processo legal; b) Regras e políticas compreensíveis; c) Competência Cultural; d) Envolvimento do Estado na Moderação de Conteúdo; e) Integridade e Explicação. Quanto aos princípios operacionais, foram mantidos os princípios de transparência, da notificação e do recurso (SANTA CLARA PRINCIPLES, 2018).

Por fim, na versão 2.0 dos Princípios de Santa Clara, estabeleceu dois princípios em relação ao Poder Público, considerando os compromissos internacionais, como por exemplo o respeito à liberdade de expressão, consagrado no art. 19 da Declaração Universal dos Direitos Humanos. São eles: 1) Removendo Barreiras à Transparência da Empresa, 2) Promovendo a Transparência do Governo (SANTA CLARA PRINCIPLES, 2018).

Ainda, temos outras cartas de tamanha importância, como a *Change the terms*, recomendações corporativas para impedir o crescimento das atividades extremistas e odiosas (CHANGE THE TERMS, 2021) e, no combate ao terrorismo, a *Christchurch Call*, elaborada após o ataque terrorista de 2019, realizado por um supremacista branco na cidade de *Christchurch*, na Nova Zelândia e transmitido ao vivo pelo *Facebook* (CRISTCHURCH CALL, 2019).

Como iniciativa individual, em 2020, o *Facebook* publicou a *Charting a Way Forward: Online Content Regulation*. Os principais pontos são: a) Como moderar conteúdo prejudicial e ao mesmo tempo preservar a liberdade de expressão? b) Como as regulamentações devem aumentar a responsabilidade das plataformas da Internet? c) A regulamentação deve exigir que as empresas de Internet cumpram certas metas de desempenho? e) A regulamentação deve definir qual “conteúdo prejudicial” deve ser proibido na internet? (FACEBOOK, 2020).

Apesar do tema moderação de conteúdo dos demais tópicos (discurso de ódio, racismo, homofobia, violência, terrorismo), especificamente no que se refere à desinformação, ganha destaque a Chamada de Paris, que a seguir será exposta.

### **2.3. Paris Call – Combate à desinformação**

A *Paris Call*, uma declaração não vinculativa com o objetivo de combater a desinformação. Ela pressupõe a importância do ciberespaço em todos os aspectos da vida, entendendo que esse deve ser aberto, seguro, estável, acessível e pacífico. A Chamada de Paris menciona ainda que o direito internacional, incluindo a Carta das Nações Unidas, o direito internacional humanitário e o direito internacional consuetudinário devem ser aplicados ao uso de tecnologias de informação e comunicação (TIC) pelos Estados (CHAMADA DE PARIS, 2018).

Ainda, a Chamada de Paris estabelece como premissa que os mesmos direitos que as pessoas têm enquanto estão *off-line* também devem ser resguardados enquanto estão *online*, reafirmando a aplicabilidade das leis internacionais de direitos humanos no ciberespaço. A chamada parte de nove princípios: a) Proteção dos indivíduos e da infraestrutura; b) Proteção da Internet; c) Defesa dos processos eleitorais; d) Defesa da propriedade intelectual; e) Não proliferação; f) Segurança do ciclo de vida; g) Higiene cibernética; h) Proteção contra o *hack* privado; i) Normas internacionais (CHAMADA DE PARIS, 2018).

O terceiro princípio, ainda que afeto diretamente ao processo eleitoral, trouxe elementos para o combate à desinformação no ciberespaço, sendo que a partir da chamada, foi criada a

Comissão Transatlântica de Integridade Eleitoral (TCEI), formada por pessoas de diferentes origens com um objetivo comum: “garantir que as pessoas decidam livremente, com base em informações independentes, quem deve representá-las.” (CHAMADA DE PARIS, 2018). A TCEI é uma iniciativa da *Alliance of Democracies Foundation*, fundada por Anders Fogh Rasmussen em 2017, numa parceria entre a Microsoft e o Governo do Canadá (CHAMADA DE PARIS, 2018).

Como parte dessas iniciativas, o Governo do Canadá realizou seis workshops, com o objetivo de evitar a interferência no processo eleitoral (CHAMADA DE PARIS, 2018). Foram trabalhados temas como a melhora do compartilhamento de informações, interferência estrangeira *versus* influência aceitável do Estado-nação e interferência eleitoral em um ambiente de pandemia. Ainda, trouxe a corresponsabilidade do combate à desinformação entre o governo, a mídia de notícias, as plataformas de mídia social, a academia e a sociedade civil, com atenção para a ameaça de interferência na infraestrutura eleitoral. Ressaltou ainda necessidade de capacitação dos cidadãos, de forma a treinar a comunidade para combater as ameaças de interferência eleitoral (GOVERNMENT OF CANADA, 2020).

Em suma, a Carta de Paris e a TCEI oferecem diversas diretrizes para o combate à desinformação, no caso voltada para o sistema eleitoral, mas que podem ser aproveitadas, em boa medida, para uma salutar moderação de conteúdo no tema da desinformação geral.

Além das cartas de princípios, surgiram iniciativas governamentais para a moderação de conteúdo, a partir de demandas específicas e históricas.

#### **2.4. A NetzDG (*Network Enforcement Act*) alemã; O Relatório Francês (*Creating a French framework to make social media platforms more accountable*) e o *Online Harms White Paper*, do Reino Unido**

O tema passou a ficar ainda mais complexo quando, em 2017, o Congresso Alemão editou a NetzDG, exigindo que as plataformas com mais de 2 milhões de usuários registrados na República Federal da Alemanha, bloqueiem ou excluam os conteúdos ilegais ou “manifestamente ilegais” sob pena de multa no importe de U\$ 50 milhões de euros (GERMAN LAW ARCHIVE, 2017).

Além da obrigação de fornecimento de relatórios detalhados das reclamações, a lei alemã prevê que o provedor facilite os canais de reclamações e que remova ou bloqueie o acesso, em 24 horas, do conteúdo “manifestamente ilegal”, após o recebimento da reclamação. Ainda, através do mesmo procedimento, que seja removido ou bloqueado o conteúdo “ilegal” em até 7 dias após a notificação. Muito embora o bom propósito de combater o discurso de ódio

e a desinformação, a NetzDG trouxe um conceito abstrato de ilegalidade, levantando questões de liberdade de expressão, uma vez que delega uma enorme quantidade de decisões a empresas privadas, com penalidades e prazos muito curtos. Vislumbra-se, portanto, que o excesso de regulamentação pode ser nocivo e até mesmo ineficaz (JORGENSEN, 2021, p. 2).

De sua vez, a França apresentou em 2019 um relatório intitulado *Creating a French framework to make social media platforms more accountable: Acting in France with a European vision* (FRENCH FRAMWORK, 2019).

De acordo com o relatório francês, a regulação deveria se basear em uma política regulatória pública que garanta as liberdades individuais e a liberdade empresarial das plataformas. Ainda, sustenta a regulamentação com foco na responsabilização das redes sociais, implementado por uma autoridade administrativa independente, com base em três pressupostos: a) transparência da função de ordenação de conteúdo; b) transparência da função que implementa os Termos de Serviço e a moderação de conteúdo; c) Defesa da integridade dos usuários. (FRENCH FRAMWORK, 2019)

O modelo francês prevê o diálogo entre os operadores, o governo, o parlamento e a sociedade civil, bem como uma cooperação europeia, reforçando a capacidade dos Estados de agir contra plataformas globais. Possui como foco, ainda, o dever de transparência das regras do jogo (FRENCH FRAMWORK, 2019).

Na mesma linha, em 2020, o Reino Unido publicou o White Paper “*Online Harms White Paper: Full government response to the consultation*”, que prevê o dever de transparências das redes, assim como um órgão independente e público, nos moldes do que constou do relatório francês (UNITED KINGDOM GOVERNMENT, 2020).

Com toda a discussão, com a pressão da sociedade civil e do governo, as plataformas optaram por definir medidas de autorregulação, a partir de uma perspectiva global, considerando que operam em diversos países.

### **3. MODERAÇÃO DE CONTEÚDO PELAS MÍDIAS SOCIAIS POR VEICULAÇÃO DE DESINFORMAÇÃO**

Diante da adesão crescente da população (usuários) da pressão externa, a atuação das mídias sociais, apesar do seu caráter privado, convive com uma série de repercussões na esfera coletiva. Há ainda um outro componente. As maiores mídias sociais operam em centenas de países, de forma que a definição das regras de convivência a partir das políticas da comunidade de cada plataforma deve ser inspirada num direito universal, reconhecido pela maioria dos

países. Natural, portanto, que os *standards* de direitos humanos fossem efeitos para inspirar a autorregulação. A escolha possui vários motivos, dentre eles a orientação no direito internacional, garantindo um padrão em vários países. Ao invés de discutir num primeiro plano a responsabilidade pelo conteúdo, considera-se inicialmente a proteção dos direitos e liberdades individuais. A partir desses pontos de convergências, as normas de freio aos Estados e empresas possui condições de seguir um mesmo padrão (JORGENSEN, 2021, p. 2)

Portanto, para estudar o funcionamento e as implicações das mídias sociais no Direito e na sociedade, exige-se a análise de questões como a responsabilidade das plataformas pelo conteúdo, a moderação, os padrões da comunidade e as políticas corporativas de direitos humanos. O ponto central é que as regras devem ser bastante claras e específica para os usuários e prever a transparência quanto ao conteúdo que será descartado, sendo que a restrição deve ser pautada nos princípios de direitos humanos. A ideia de visão social dos princípios de direitos humanos, garante a inclusão e os pontos de vista diferentes, prevê restrições ao conteúdo à violência, ódio e desinformação, garantindo como consequência uma proteção individual (JORGENSEN, 2021, p. 3).

Para exemplificar, deve-se perceber que o Brasil, que possui o sistema jurídico baseado nas tradições do direito civil, trazendo a lei como fonte primária, trataria o tema de forma diversa de muitos países seguem a tradição da common law, ressaltando-se assim a importância da utilização do direito internacional, que contém elementos de ambos os sistemas. (PINTO; ALVEZ, 2020, p. 82)

No Brasil, aliás, ressalta-se a existência do Comitê Gestor da Internet – CGI, criado pela Portaria Interministerial nº 147/1995, posteriormente alterada pelo Decreto nº 4.829/2003. O Comitê é formado por órgãos estatais, representante privados, do terceiro setor e da sociedade civil organizada, tendo como atribuições, dentre outras, estabelecer diretrizes estratégica relacionadas ao uso e desenvolvimento da Internet no Brasil. A Resolução nº 2009/003, fixou os princípios para a governança e o uso da *internet* no Brasil, destacando-se dentre outros aspectos a liberdade, privacidade e os direitos humanos (COMITÊ GESTOR DA INTERNET NO BRASIL, 2009).

Importante mencionar, ainda no âmbito nacional, o Marco Civil da *Internet*, trazido pela Lei 12.965, de 23 de abril de 2014, que estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Referido diploma legal tem como fundamento o direito à liberdade de expressão, bem como os direitos humanos, a livre iniciativa, a livre concorrência, a finalidade social da rede, dentre outros não menos importantes.

Guarda relevância para o tema do presente trabalho a Seção III do Marco Civil da *Internet*, que trata da responsabilidade pelos conteúdos gerados por terceiros, e que estabelece no seu artigo 18 que o provedor de conexão à internet não será responsabilizado civilmente. Muito embora o conceito de mídias sociais ser distinto do conceito provedor de conexão à internet, fica claro, ainda que por analogia, que o legislador pátrio fez a opção pela não responsabilização, a não ser que se descumpra ordem judicial, referida no artigo 19.

Merece atenção especial o artigo 21 do Marco Regulatório da *Internet*, que prevê a responsabilidade subsidiária do “provedor de aplicações de internet” pela violação da intimidade, sem autorização de seus participantes, por conteúdo contendo “cenas de nudez ou de atos sexuais de caráter privado”, após a notificação do interessado. O mesmo artigo aponta uma outra figura, qual seja, “provedor de aplicações de internet”, o que se aproxima do conceito de mídias sociais.

Mas como se verifica, a preocupação originária do legislador nacional foi centrada em apenas um dos tópicos que merecer a atenção das redes, da sociedade e do governo, as cenas de nudez e atos sexuais privados.

Retomando ao tema em escala universal, deve ser considerado que o modelo de negócio das mídias sociais é bem diferente da indústria tradicional. No caso, os usuários das plataformas de mídias sociais não compram o produto, eles pagam pelo acesso com seus dados, que são transformados em receita. Portanto, a proteção ao usuário ganha relevo e precisa ser objeto de atenção especial, com o objetivo de não causar danos (BUHMANN; OLIVERA, 2020, p. 14).

O esforço centrado na autorregulação fez o grupo *Meta*, detentor do *Facebook*, *Instagram* e *WhatsApp*, além de suas políticas de privacidade e termos de uso, editar a sua política de Direitos Humanos, disponibilizando em seu *site* oficial, com objetivo de informar seus usuários, a sua Política Corporativa de Direitos Humanos (FACEBOOK, 2021).

De igual forma, o *Youtube*, que pertence à empresa *Google*, além de suas políticas internas, segue os padrões de Direitos Humanos da *Google* reconhecidos internacionalmente, disponíveis na página oficial (GOOGLE, 2020).

O *Twitter* também possui e disponibiliza em seu *site* sua Política Pública para Defesa e Respeito aos direitos das pessoas que utilizam seus serviços. Na política, o *Twitter* enfatiza seu compromisso com a liberdade de expressão e a privacidade. A Política de Direitos Humanos do *Twitter* tem como base a Declaração de Direitos dos Estados Unidos e a Convenção Europeia de Direitos Humanos, que objetiva a Proteção dos Direitos Humanos de Liberdade e, também, os Princípios Orientadores sobre Empresas e Direitos Humanos das Nações Unidas (PO's). (TWITTER)

Para exemplificar a complexidade e o resultado dessa abordagem calçada nos direitos humanos, o *Facebook*, maior mídia social do mundo, possui seus *standards* da comunidade inspirados nos preceitos universais, com a ressalva de que diante das milhares de postagens por segundo, é inegável que só existe uma possibilidade para o primeiro filtro: a inteligência artificial. Sander (2021, p. 163) assinala que para cumprir a tarefa de moderar o conteúdo, as mídias sociais contam com a inteligência artificial a partir dos algoritmos e da revisão humana, a depender do tamanho, dos recursos e da cultura da plataforma.

Os padrões da comunidade do *Facebook* (*standards*) pretendem garantir a autenticidade, segurança, privacidade e dignidade. No item integridade/autenticidade, se encontra o tópico desinformação (FACEBOOK, 2018).

No tópico desinformação, no qual vem mencionada a dificuldade de definir a notícia falsa, o esforço da rede para reduzir a disseminação de boatos e a desinformação viral, além de direcionar usuários para informações oficiais, informar a proibição de contas falsas, a fraude e o comportamento inautêntico coordenado (FACEBOOK, 2018).

No mesmo campo, o *Facebook* informa as desinformações que serão removidas: a) desinformação com risco de agressão física ou violência; b) desinformação prejudicial sobre a saúde; c) desinformação que pode contribuir ou contribua diretamente para o risco de interferência na capacidade das pessoas participarem do pleito eleitoral; d) mídia manipulada, que pode se tornar viral rapidamente (FACEBOOK, 2018).

O *Facebook* reconhece, portanto, a dificuldade de definir o que seria desinformação, sobretudo para que não se invada o direito à liberdade de expressão. Contudo, elegeu meios para o combate, através da disseminação da informação oficial e confiável e do combate à fraude e contas falsas. Ainda, de acordo com seus princípios, informou que removerá desinformação na área da saúde, pleito eleitoral, que traz risco à agressão ou violência, assim como a que viraliza rapidamente.

Os esforços das redes, devem ainda, considerar que na moderação de conteúdo, uma concepção estrutural dos direitos humanos deve se dar a partir de uma visão holística, assim como uma abordagem cultural daria maior ênfase na obrigação de proteger a liberdade de expressão, como base para exigir que os Estados assegurem mecanismos efetivos de transparência, processo legal, responsabilidade e supervisão da plataforma, numa relação público-privada (SANDER, 2021, p. 192).

Portanto, além do combate à desinformação, deve ser garantida a transparência das redes, a facilitação dos meios de revisão das decisões, regras claras quanto ao que é considerado desinformação, assim como o respeito às diferenças culturais e línguas locais.

#### 4. CONCLUSÃO

O estudo demonstrou que é inegável que o avanço exponencial da tecnologia da informação e comunicação, bem como com a democratização da *internet*. Nesse contexto, o problema da desinformação foi acentuado a partir caso da *Cambridge Analytica*, em que o garimpo e tratamento de dados, aplicados no convencimento individual do eleitor, mudou os rumos decisórios e atingiu princípios basilares da democracia.

Apesar das repercussões na esfera coletiva, tratar do tema implica também reconhecer o caráter privado das mídias sociais, sua necessidade de sobrevivência no mercado, tendo como base o interesse na segurança do usuário e a liberdade de expressão.

A partir da análise do itinerário histórico, com a abordagem dos primeiros casos judiciais, compreende-se que a implicação da responsabilização do intermediário que se dispunha a moderar o conteúdo, fez com que as plataformas, criadas e desenvolvidas a partir desse contexto, foram tomadas pelo dilema de que moderar seria um risco.

Após, por pressão governamental, da sociedade e do mercado, as mídias sociais se viram obrigadas a moderar conteúdo, sobretudo nas áreas mais sensíveis. Nesse contexto, importante compreender a importância dos *standards* de direitos humanos, por serem universais e com a possibilidade de aplicação em centenas de países. Assim, as mídias sociais inspiraram seus Padrões da Comunidade nos *standards* dos direitos humanos, como garantia de padronização frente a legislação interna dos países em que operam. Ou seja, a tarefa que já é complexa, seria praticamente impossível não fossem os *standards* de direitos humanos.

Por meio da valoração dos princípios que regem os direitos humanos, encontrou-se uma solução para equacionar a aparente dicotomia entre a liberdade de expressão e a necessidade de garantia de segurança nas redes diante da proliferação da desinformação, do discurso de ódio, da violência, da exploração sexual, dentre outras práticas nocivas para o indivíduo (usuário) e para a sociedade.

Por fim, deve ser considerado que as cartas de princípios apontam diversas diretrizes que, somadas aos *standards* de direitos humanos, dão um horizonte mais alvissareiro para o aprimoramento da moderação de conteúdo. O dever de transparência, definindo parâmetros claros e divulgação de dados, o reconhecimento das particularidades culturais e da língua local, a possibilidade de revisão (recurso), dentre outras medidas, precisam ser implementadas em conjunto com os padrões de direitos humanos, formando um conjunto de medidas para o combate à desinformação por meio da moderação de conteúdo.

## REFERÊNCIAS

AIETA, Vania Siciliano. **O Impacto Eleitoral Resultante da Manipulação das Fake News no Universo das Redes Sociais: a Construção da Desinformação**. Revista Interdisciplinar de Direito. v. 18, n. 1, pp 213-233, jan./jun. 2020.

BRASIL. *Lei n. 12.965, de 23 de abril de 2014. Marco Regulatório da Internet*. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/112965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm). Acesso em: 15 abr 2022.

CARVALHO, Lucas Borges de. A democracia frustrada: fake news, política e liberdade de expressão nas redes sociais. **Internet&sociedade**. São Paulo, SP, v. 1, n. 1, p. 172-199, fev. 2020. Disponível em: <[https://revista.internetlab.org.br/wp-content/uploads/2020/02/ilab.01.revista01\\_0214-B-arrastado-2.pdf](https://revista.internetlab.org.br/wp-content/uploads/2020/02/ilab.01.revista01_0214-B-arrastado-2.pdf)> Acesso em: 17 abr. 2022

COMITÊ GESTOR DA INTERNET NO BRASIL. **Resolução nº 2009/003**. Disponível em <https://www.cgi.br/resolucoes/documento/2009/003/>. Acesso em: 14 mar. 2022.

CHANGE THE TERMS. **Reduzir o ódio e a desinformação on line**. 2021. Disponível em <https://www.changetheterms.org/> Acesso em: 01 mar. 2022.

CHIRWA, Candice and MANYANA, Zimkhitha. **The Rise of Fake News: Surveying the Effects of Social Media on Informed Democracy**. The Thinker, 2021.

THE CENTER FOR REGULATORY EFFECTIVENESS. **Creating a French Framework to make social media platforms more accountable: acting in France with a European vision**. Disponível em <https://thecre.com/RegSM/wp-content/uploads/2019/05/French-Framework-for-Social-Media-Platforms.pdf> Acesso em: 01 mar. 2022

CRISTCHURCH CALL. **Carta de Cristchurch**. Disponível em <https://www.christchurchcall.com/> Acesso em: 01 mar. 2022.

D'ANCONA, Matthew. **Pós-verdade, a nova guerra contra os fatos em tempos de fake news**. São Paulo: Faro Editorial, 2018.

EMPOLI, Giuliano Da. **Os engenheiros do caos**. São Paulo: Vestígio, 2019.

ESTARQUE, Marina; ACHEGAS, João Victor; BOTTINO, Celina; PERRONE, Christian. **Redes Sociais e Moderação de Conteúdo: criando regras para o debate público a partir da esfera privada**. Instituto de Tecnologia e Sociedade do Rio – ITS. Disponível em [https://itsrio.org/wp-content/uploads/2021/04/Relatorio\\_RedetesSociaisModeracaoDeConteudo.pdf](https://itsrio.org/wp-content/uploads/2021/04/Relatorio_RedetesSociaisModeracaoDeConteudo.pdf). Acesso em: 27 fev. 2022.

FACEBOOK. **Charting a way foward: online content regulations**. Disponível em [https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward\\_Online-Content-Regulation-White-Paper-1.pdf](https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf) Acesso em: 01 mar. 2022.

FACEBOOK. *Corporate Human Rights Policy*, 2021. Disponível em <https://about.fb.com/wp-content/uploads/2021/03/Facebooks-Corporate-Human-Rights-Policy.pdf>. Acesso em: 17 abr. 2022.

FACEBOOK. **Padrões da Comunidade do Facebook**, 2018. Disponível em <https://transparency.fb.com/pt-br/policies/community-standards/?source=https%3A%2F%2Fwww.facebook.com%2Fcommunitystandards%2F> Acesso em: 17 abr. 2022.

FORNASIER, Mateus de Oliveira; BECK, Cesar. (2020). **Cambridge analytica: escândalo, legado e possíveis futuros para a democracia**. *Revista Direito Em Debate*, 29(53), 182-195. Disponível em: < <https://doi.org/10.21527/2176-6622.2020.53.182-195> >. Acesso em: 17 abr 2022.

GOOGLE. *Human Rights*. 2020. Disponível em <https://about.google/intl/pt-BR/human-rights/> Acesso em: 17 abr. 2022.

GOVERNMENT OF CANADA. *Paris call for trust and security in cyberspace*. Disponível em <https://www.canada.ca/en/democratic-institutions/services/paris-call-trust-security-cyberspace.html>. Acesso em: 11 mar. 2022.

H2O. *Stratton Oakmont, inc. v Prodigy Services Co.* Disponível em <https://h2o.law.harvard.edu/cases/4540> Acesso em: 27 fev. 2022.

IZA, Daniela Gonzales, **Human Rights In The Digital Era: Challenges and Opportunities From The United Nations Human Rights System**. *Jurnal Pengajian Umum Asia Tenggara* 22 (2021): 1 – 13

JORGENSEN, Rikke Frank. **A Human Rights-Based Approach to Social Media Platforms: Is Social Media Ethical?**. 2021. Disponível em <https://berkleycenter.georgetown.edu/responses/a-human-rights-based-approach-to-social-media-platforms>. Acesso em: 16 mar. 2022.

JUSTIA US LAW. **Cubby, Inc. v. CompuServe Inc., 776 F. Supp. 135 (SDNY 1991)**. Disponível em <https://law.justia.com/cases/federal/district-courts/FSupp/776/135/2340509/> Acessado em 27 fev. 2022

LEGAL INFORMATION INSTITUTE. **Primeira Emenda**. Disponível em [https://www.law.cornell.edu/constitution/first\\_amendment](https://www.law.cornell.edu/constitution/first_amendment) Acesso em: 01 mar. 2022.

MANILA PRINCIPLES. **Princípios de Manila**. Disponível em <https://manilaprinciples.org/pt-br.html> Acesso em: 01 mar. 2022.

NEISSER, Fernando Gaspar. **Crime e Mentira na Política**. Belo Horizonte: Editora Fórum, 2016.

PARIS CALL. **Carta de Paris**. Disponível em <https://pariscall.international/en/> Acesso em: 01 mar. 2022.

PINTO, Thiago Alves; ALVES, **Rodrigo Vitorino Souza**. **Investigations on the Use of Limitations to Freedom of Religion or Belief in Brazil**. *Religion and Human Rights* 15 (2020) 77–95

RAIS, Diogo ... [et al.]. **Direito Eleitoral Digital**. São Paulo: Thompson Reuters Brasil, 2018.

REDES SOCIAIS NAS ELEIÇÕES 2018. FGV. **Sala de Democracia Digital**. Disponível em: <https://bibliotecadigital.fgv.br/dspace/handle/10438/25737>. Acesso em: 17 abr 2022.

SANDER, Barrie. **Democratic Disruption in the Age of Social Media: Between Marketized and Structural Conceptions of Human Rights Law**. *The European Journal of International Law* Vol. 32 no. 1, 2021.

SANTA CLARA PRINCIPLES. **Os princípios de Santa Clara**. Disponível em <https://santaclaraprinciples.org/pt/cfp/> Acesso em: 01 mar. 2022.

TANDOC JR., Edson C.; LIM, Zheng Wei; LING, Richard. *Defining “Fake News”: A typology of acholarly definitions*. **Digital Journalism**, Londres, v.6, n.2, p.137-153, 30 ago. 2017. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1360143?scroll=top&needAccess=true>. Acesso em: 17 abr 2022.

TWITTER. **Defender e respeitar as pessoas que utilizam os nossos serviços**. Disponível em <https://help.twitter.com/en/rules-and-policies/defending-and-respecting-our-users-voice> Acesso em: 17 de abr 2022.

UNITED KINGDOM GOVERNMENT. **Online Harms White Paper: Full government response to the consultation**. Disponível em <https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response>. Acesso em: 01 mar. 2022.