

Uma Nova Cara para a Política? Considerações sobre *Deepfakes* e Democracia

A New Face for Politics? Considerations on Deepfakes and Democracy

CAITLIN MULHOLLAND¹

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Rio de Janeiro (RJ). Brasil.

SAMUEL RODRIGUES DE OLIVEIRA²

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Rio de Janeiro (RJ). Brasil.

RESUMO: *Deepfakes* – mídias audiovisuais hiper-realistas, manipuladas digitalmente mediante técnicas de inteligência artificial – são um fenômeno recente e, por isso, carecedor de debate. O objetivo deste artigo é investigar a influência que *deepfakes* exercem (ou podem exercer) no debate público e, consequentemente, no processo político como um todo. Para tanto, em um primeiro momento, realizamos uma construção teórica acerca da mediatização da política, da influência de fofocas e escândalos no processo democrático e da polarização da “esfera pública conectada”, analisando como esses fenômenos corroboram aquilo que se denomina crise da democracia representativa. Em seguida, apresentamos e discutimos com maior detalhamento as chamadas *deepfakes*, explorando a sua relação com o fenômeno da desinformação, como elas se inserem no contexto do debate político na dita “Era da Informação” e como afetam (ou podem afetar) esse momento do processo democrático. Concluímos que, embora sejam um fenômeno relativamente novo e merecedor de atenção, as *deepfakes*, ao menos no que se refere ao processo político, não fazem muito além de aprofundar determinados problemas já existentes em nossa sociedade.

PALAVRAS-CHAVE: Crise democrática; *deepfake*; *fake news*; inteligência artificial; Internet; polarização.

ABSTRACT: Deepfakes – hyper-realistic audiovisual media, digitally manipulated through artificial intelligence techniques – are a recent phenomenon and, therefore, worthy of debate. The goal of this paper is to investigate the influence deepfakes have (or may have) on public debate and, consequently, on the political process overall. For this purpose, in a first moment, we perform a theoretical construction about the mediatization of politics, the influence of gossip and scandals in

1 Orcid: <http://orcid.org/0000-0002-0951-3158>.

2 Orcid: <http://orcid.org/0000-0002-7059-580X>.

the democratic process, and the polarization of the “connected public sphere”, analyzing how these phenomena corroborate what is called the crisis of representative democracy. Next, we present and discuss in more detail the so-called deepfakes, exploring their relationship with the phenomenon of misinformation, how they fit into the context of political debate in the so-called “information age”, and how they affect (or may affect) this moment of the democratic process. We conclude that, although deepfakes are a relatively new phenomenon and deserving of attention, at least as far as the political process is concerned, they do little more than deepen certain problems already in existence in our society.

KEYWORDS: Artificial intelligence; crisis of democracy; deepfake; fake news; Internet; polarization.

SUMÁRIO: Introdução: quando os olhos deixam de ser testemunhas confiáveis; 1 Mediatização da política, polarização e crise democrática; 1.1 (Des)mediatização da política e política midiática; 1.2 Fofocas, escândalos e política: da polarização à crise da democracia; 2 *Deepfakes*: catalisadores da ruptura?; 2.1 Afinal, o que são *deepfakes*?; 2.2 *Deepfakes*, erosão da confiabilidade e o “dividendo do mentiroso”; Considerações finais: nem tudo está perdido; Referências.

INTRODUÇÃO: QUANDO OS OLHOS DEIXAM DE SER TESTEMUNHAS CONFIÁVEIS

Estamos entrando numa era em que nossos inimigos podem fazer parecer que qualquer um está dizendo qualquer coisa em qualquer momento, mesmo que nunca tenham dito essas coisas. Por exemplo, eles poderiam me fazer dizer coisas como [...] “o Presidente Trump é um total e completo imbecil”³.

Essas são as palavras ditas por Barack Obama, ex-presidente dos Estados Unidos da América, em um vídeo recente, publicado no YouTube. Quer dizer, não exatamente por Obama. O rosto que vemos e a voz que ouvimos no vídeo são, de fato, do ex-presidente, e os movimentos realizados parecem verdadeiros. Contudo, trata-se de uma *deepfake*⁴ – um vídeo hiper-realista cujo conteúdo fora digitalmente manipulado.

Deepfakes são um fenômeno recente – as primeiras manipulações digitais hiper-realistas que receberam esse nome remetem ao ano de 2017 (Kietzmann et al., 2020) – e de grande ubiquidade (Citron; Chesney, 2019). Os seus usos variam desde a criação de vídeos humorísticos e conteúdos educacionais até a falsificação de vídeos de conteúdo pornográfico, ge-

3 Cf.: “You Won’t Believe What Obama Says In This Video!”. Disponível em: <https://www.youtube.com/watch?v=cQ54GDm1eL0&t=8s>. Acesso em: 21 jan. 2021.

4 O termo poderia ser traduzido, ainda que com perda de sentido, como “falsificação profunda”.

almente utilizados para fins de *revenge porn*⁵, e a manipulação de falas de políticos e celebridades. Ainda que traga consigo diversos benefícios⁶, inúmeros também podem ser os riscos e as ameaças postos por essa nova tecnologia.

A maneira como nos comunicamos vem, ao longo dos últimos anos, se transformando e evoluindo em velocidade e sentidos jamais vistos: como a comunicação é a base das atividades humanas, “todos os domínios da vida social estão sendo modificados pelos usos disseminados da Internet” (Castells, 2003). Atualmente, o consumo de informações por meio de vídeos, especialmente em dispositivos móveis, encontra-se em uma curva de ascensão exponencial (Teyssou; Spangenberg, 2019), e é justamente por isso que as *deepfakes* tanto nos importam. A digitalização da informação, somada ao surgimento de redes sociais, resultou em mudanças fundamentais no que diz respeito à coleta e difusão de informações: estatísticas do YouTube, referentes a outubro de 2018, apontaram a existência de mais de um bilhão de usuários da plataforma de vídeo, com um bilhão de horas assistidas diariamente, 70% delas em dispositivos móveis (Teyssou; Spangenberg, 2019). Hoje, um a cada cinco usuários da Internet se informa através da plataforma, que fica atrás apenas do Facebook no que diz respeito ao número de usuários (Westerlund, 2020).

A atração exercida pelos meios de comunicação audiovisual pode ser explicada por aquilo que os psicólogos denominam “fluência de processamento”, que corresponde ao nosso viés cognitivo inconsciente de predileção por informações as quais nosso cérebro pode processar rapidamente (Schick, 2020). Isso pode ser problemático, como veremos a seguir, porque

5 O termo pode ser traduzido como “pornografia de vingança”. Sobre o tema, cf.: HAO, Karen. An AI app that “undressed” women shows how deepfakes harm the most vulnerable. *MIT Technology Review*, [s.l.], 28 jun. 2019a. Disponível em: <https://www.technologyreview.com/2019/06/28/134352/an-ai-app-that-undressed-women-shows-how-deepfakes-harm-the-most-vulnerable/>. Acesso em: 29 jul. 2020.

6 *Deepfakes* podem ser utilizadas em diversos sentidos e para diferentes fins: desde objetivos meramente comerciais, como para as indústrias do entretenimento, videogames, moda e *e-commerce*, passando pelas áreas das artes e da educação, chegando à esfera de construção da subjetividade e autonomia pessoal. A título exemplificativo, espera-se que, futuramente, pessoas com Alzheimer que tenham dificuldade de reconhecer parentes ou amigos poderão interagir com rostos dos quais ainda se lembram, graças a manipulações digitais que realizem o seu rejuvenescimento (cf.: “Terrifying high-tech porn: Creepy ‘deepfake’ videos are on the rise”. Disponível em: <https://www.foxnews.com/tech/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-the-rise>. Acesso em: 18 jun. 2021). Outro interessante exemplo, que já se encontra em uso, é a possibilidade de restauração da capacidade de fala de pessoas com esclerose lateral amiotrófica (ELA) por meio de tecnologias de edição de áudio, de maneira semelhante ao que fez o astrofísico Stephen Hawking (cf.: “Lyrebird Helps ALS Ice Bucket Challenge Co-Founder Pat Quinn Get His Voice Back”. Disponível em: https://www.huffingtonpost.ca/2018/04/14/lyrebird-helps-als-ice-bucket-challenge-co-founder-pat-quinn-get-his-voice-back_a_23411403/. Acesso em: 18 jun. 2021).

as pessoas tendem a acreditar no que veem, de modo que imagens e outras formas de mídias digitais são frequentemente aceitas enquanto verdadeiras sem questionamentos (Maras; Alexandrou, 2019). Como nós somos, em certo sentido, “programados” para acreditar naquilo que vemos ou ouvimos, temos a tendência de considerar vídeos e áudios como autênticos e incorruptíveis.

Embora as *deepfakes* suscitem inúmeras discussões tão interessantes quanto urgentes, é preciso que direcionemos nossa atenção a um ponto específico. Considerando que a Internet tem sido um dos principais (senão o principal) meios de participação popular nos processos políticos (Magrani, 2014), e que as *deepfakes* são mídias amplamente difundidas sobretudo em redes sociais, buscaremos analisar como o seu uso afeta, ou pode afetar, o debate público e, conseqüentemente, o processo de disputa política. A hipótese que trazemos é que, quando usadas para fins políticos, essas falsificações digitais podem manipular de maneira significativa a opinião pública e minar a confiança dos cidadãos nas instituições políticas, corroborando o enfraquecimento da própria democracia. Nossa intenção é demonstrar como (ou se) a participação popular nos processos políticos pode ser afetada por essa inovação tecnológica, e, com isso, chamar a atenção para a necessidade de se discutir com maior atenção a temática.

A metodologia aplicada à pesquisa para o desenvolvimento do artigo consiste em análise teórica crítica, embasada em investigação bibliográfica interdisciplinar nas áreas do direito, sociologia, ética e tecnologia (especialmente nas áreas de inteligência artificial, algoritmos, *big data* e *machine learning*). O método empregado é o dedutivo, uma vez que partimos da análise estrutural de um contexto – a mediatização da política e o processo de polarização e crise decorrente da política midiática – para o estudo de um aspecto específico deste contexto – a possível influência das *deepfakes* no debate político.

Para tanto, esta pesquisa encontra-se dividida em dois momentos principais. No primeiro deles, discutimos a noção de (des)mediatização da política e política midiática, sobretudo a partir de Manuel Castells (2009, 2018a), Jeanette Hofmann (2019) e Byung-Chul Han (2017, 2018a, 2018b). Ainda no primeiro tópico, investigamos, novamente a partir de Castells (2009), mas também a partir de Solove (2007) e Han (2017), o processo de escandalização da política; com base em Cass Sunstein (2007, 2018), principalmente, discutimos a (ciber)polarização política, que corrobora a crise da democracia. Em um segundo momento, apresentamos e debatemos

as *deepfakes* a partir de autores como Danielle Citron e Bobby Chesney (2019). Discutimos as implicações dessa nova tecnologia na sociedade de informação a partir da contribuição de Nina Schick (2020), Eli Pariser (2012) e, uma vez mais, Sunstein (2018). Buscamos, ao final, “amarrar” todos os conceitos e premissas apresentados ao longo do texto, apresentando breves considerações finais.

1 MEDIATIZAÇÃO DA POLÍTICA, POLARIZAÇÃO E CRISE DEMOCRÁTICA

1.1 (DES)MEDIATIZAÇÃO DA POLÍTICA E POLÍTICA MIDIÁTICA

De acordo com Jeanette Hofmann (2019), novas tecnologias de comunicação digital deram origem a novos contornos do desenvolvimento dos meios de comunicação, os quais refletem as transformações estruturais mais amplas das sociedades ocidentais. Para a autora, uma vez que a democracia representativa requer aparatos de mediação técnica, a democracia moderna e os meios de comunicação estão estreitamente interligados: não há que se falar em democracia mediatizada como um tipo de democracia, senão apenas como uma perspectiva para se enxergar os processos democráticos atuais (Hofmann, 2019). Formas específicas de comunicação, bem como novos tipos de mídia, surgem em conjunto com processos sociais ainda mais complexos, cocriando-se e possibilitando mutuamente suas existências (Hofmann, 2019, p. 9).

Stefano Rodotà se refere a esse fenômeno como “tecnopolítica”. Nessa conjuntura, as tecnologias de informação e comunicação exercem influência direta na esfera política, em diversos aspectos – desde o aumento da participação do cidadão em todos os momentos da democracia até a possibilidade de se cobrar maior transparência dos governantes (Rodotà, 2008, p. 91). É nesse mesmo sentido que Yochai Benkler afirma que vivemos em uma “esfera pública conectada”, que permite que os “indivíduos, atuando sozinhos ou com outros, sejam participantes ativos da esfera pública, em vez de leitores, ouvintes ou espectadores passivos” (Benkler, 2006, p. 212, tradução nossa)⁷. Nas palavras de Rodotà,

a passagem da comunicação vertical, típica da televisão tradicional, à comunicação horizontal, típica da internet, permitiria de fato resgatar o cidadão da

7 No original: “Individuals, acting alone or with others, to be active participants in the public sphere as opposed to its passive readers, listeners, or viewers”.

passividade de espectador, tornando-o protagonista de um processo no qual desaparecem as distinções entre produtores e consumidores de informações. (2008, p. 92)

Para Castells (2009), no contexto histórico em que vivemos, a política consiste majoritariamente na política midiática. Uma vez que os meios de comunicação constituem a nova ágora na qual relações de poder são disputadas e estabelecidas, basicamente todos os atores políticos e sociais precisam buscar seu espaço na mídia, a fim de se fazerem ouvir e alcançarem seus objetivos, pois

em virtude dos efeitos convergentes da crise dos sistemas políticos tradicionais e do grau de penetrabilidade bem maior dos novos meios de comunicação, a comunicação e as informações políticas são capturadas essencialmente no espaço da mídia. Tudo o que fica de fora do alcance da mídia assume a condição de marginalidade política. (Castells, 2018a, localização 10.944)

Em um contexto de mediatização da política, não estar presente na mídia é virtualmente equivalente a não existir na mente das pessoas, de modo que apenas as organizações e os políticos que se comunicam diretamente e em larga escala com os cidadãos possuem alguma chance efetiva de influenciarem suas decisões e, como resultado, elevarem-se às posições de poder e/ou garantir a manutenção do *status quo*, quando já se encontram nessas posições (Castells, 2009, p. 194). Embora a disputa pelo poder seja uma característica intrínseca à própria política, sobretudo em se tratando de sistemas democráticos, e embora a comunicação seja um fenômeno tão antigo quanto a humanidade em si, há um fator relativamente novo que deve ser levado em consideração quando pensamos o debate político nos dias atuais. Trata-se do surgimento da autocomunicação em massa, um processo de transformação tecnológica e organizacional da comunicação, “baseada em redes horizontais de comunicação multidirecional, interativa, na Internet; e, mais ainda, nas redes de comunicação sem fio, atualmente a principal plataforma de comunicação em toda parte” (Castells, 2013, p. 128).

Nesse sentido, um ponto a se considerar é que, há até relativamente pouco tempo, o acesso à mídia era provido e controlado por *gatekeepers* (Castells, 2009) ou intermediários (Han, 2018a), pelos detentores dos meios de comunicação tradicional – emissoras de televisão e de rádio, editores e/ou donos de jornais e revistas. Esse cenário se transformou profundamente com o advento da Internet e da chamada Era da Informação, caracterizada pela existência de redes horizontais de comunicação e o consequente sur-

gimento das redes sociais (Castells, 2013). Nas palavras de Han, “a comunicação digital se caracteriza pelo fato de que informações são produzidas, enviadas e recebidas sem mediação por meio de intermediários” (2017a, posição 267), e não mais filtradas ou dirigidas por mediadores.

Ainda de acordo com Castells, as redes sociais “conectam todas as dimensões da vida das pessoas” (Castells, 2013, p. 136), sendo essa uma tendência importante para toda a sociedade, sobretudo porque gera a transformação da cultura e das relações humanas ao estimular o compartilhamento ininterrupto de informações. Na Era da Informação, o mundo se encontra permanentemente conectado. Basicamente todas as experiências humanas passaram a ser compartilhadas na rede, o que modificou profundamente diversas esferas das nossas vidas. Entre elas, sem dúvidas, a política. Ao discutir a relevância da mídia para a comunicação humana nos tempos atuais, Castells reconhece que o poder da comunicação alterou inclusive a política democrática, *i.e.*, a política baseada em eleições livres e supostamente justas como o principal mecanismo de acesso aos cargos políticos (Castells, 2009, p. 194). Como resultado, “a mídia eletrônica (não só o rádio e a televisão, mas todas as formas de comunicação, tais como o jornal e a internet) passou a se tornar o espaço privilegiado da política” (Castells, 2018a, localização 10.925).

Curiosamente, Byung-Chul Han (2018a) dá um nome diametralmente oposto ao fenômeno que Castells denomina “mediatização”. Segundo o autor sul-coreano, as novas formas de comunicação digital se distinguem das *mídias* tradicionais em diversos aspectos, e a sua (quase) prevalência sobre os meios de comunicação tradicional “desmediatizam a comunicação” (Han, 2018a, posição 299). Como toda a sociedade de informação se apoia na comunicação desmediatizada, a desmediatização se estende também à política, ameaçando a própria democracia representativa, na medida em que os cidadãos passam a enxergar os representantes políticos não como transmissores de informação, mas sim como barreiras à manifestação de suas opiniões (Han, 2018a). Explica-se: segundo o autor, a atual busca por informação e transparência não é uma demanda política, reivindicada para os processos políticos em si, como a tomada de decisões de interesse público, senão uma tentativa de “desmascara ou expor a classe dos políticos” e “transformar indivíduos em objetos de escândalo” (Han, 2018b, p. 21).

Em que pese a distinta conceituação do fenômeno, os autores parecem concordar no que se refere a seus efeitos, que se encontram sinte-

tizados no pensamento de Hofmann. Para a autora alemã, o surgimento e, principalmente, o ímpeto da autocomunicação em massa reflete tanto as propriedades materiais dos novos meios de comunicação quanto as transformações pelas quais têm passado a democracia e a esfera pública (Hofmann, 2019, p. 10). À medida que a esfera pública global e sua infraestrutura midiática ganham importância, e que os limites econômicos e jurídicos entre a produção e o consumo de notícias se tornam “desinstitucionalizados” (Hofmann, 2019, p. 10), resta cada vez mais evidente como a Internet constitui um espaço (mais) aberto e eficaz no que se refere à democratização da comunicação e, conseqüentemente, do debate público (Magrani, 2014, p. 131).

As transformações no/do processo político são particularmente relevantes quando consideramos que, especialmente nas duas últimas décadas, o crescimento da centralidade da política midiática coincidiu com personalização da política (Castells, 2009, p. 202). O aumento da notoriedade de figuras políticas em detrimento da importância de partidos ou alianças políticas e de afinidades ideológicas tem caracterizado as disputas políticas em todo o mundo, o que é, ao mesmo tempo, causa e consequência do aumento do poder da comunicação e da política midiática: uma vez que todas as organizações midiáticas, sejam elas focadas na comunicação em massa ou na autocomunicação em massa, ou ambas, têm como objetivo expandir sua influência e seus recursos ao expandir e aumentar suas audiências (Castells, 2009, p. 195), o uso de figuras políticas carismáticas ou polêmicas se revela uma ferramenta valiosa para capturar a atenção das pessoas.

As pessoas tomam decisões, incluindo-se aquelas relacionadas à política, baseadas em informações – e sobretudo imagens – que são produzidas e divulgadas pela mídia tradicional e na Internet (Castells, 2009). As imagens com as quais temos contato geram um primeiro reflexo emocional, que é seguido de um processo cognitivo de elaboração e decisão; a impressão do contato inicial se transforma em opinião, a qual, por sua vez, “se confirma ou se desmente na elaboração do debate contínuo que acontece nas redes sociais em interação com a mídia” (Castells, 2018b, localização 253). As predisposições pessoais, que se encontram “armazenadas” nas mentes das pessoas em razão de suas experiências cotidianas, são trabalhadas pelas campanhas eleitorais (Castells, 2009, p. 224), a fim de, obviamente, conquistar o voto de tantos cidadãos quanto possível. Nesse sentido, as campanhas eleitorais são um elemento-chave para o acesso dos candidatos às instâncias de poder, desempenhando o papel de “motor da democracia”

(Castells, 2009, p. 228). Elas são ainda, junto da personalização da política, uma das características fundamentais da política midiática (Castells, 2009)⁸.

A política midiática na era digital, por sua vez, é caracterizada justamente pela interação entre a mídia tradicional e as redes sociais (Castells, 2009, p. 224). Castells, no início dos anos 2000, afirmou que os meios de comunicação “convencionais” continuam sendo as principais fontes de informação com as quais os cidadãos têm contato, ao passo que a Internet abre “pontos de entrada” para o universo da comunicação audiovisual em massa (Castells, 2009, p. 224). Todavia, a interação entre essas diferentes esferas persista, o que observamos hoje, ao menos no Brasil e em boa parte do mundo, é a predominância dos meios de autocomunicação em massa no processo de informação e formação de opinião das pessoas. Por aqui, a rede de mensagens WhatsApp conta com pelo menos 130 milhões de usuários⁹. Alarmantemente, o aplicativo é utilizado como principal fonte de informação de 79% da população brasileira, seguido do YouTube (49%) e do Facebook (44%)¹⁰.

Tais dados se tornam ainda mais preocupantes quando pensamos em uma das principais diferenças entre os meios de comunicação em massa tradicionais e as redes sociais: a capacidade de controle e filtragem da entrada de informações no ecossistema informativo, inexistente nos meios “desmediatizados” (Han, 2018a). Os filtros estabelecidos por proprietários, anunciantes, editores e jornalistas profissionais privilegiam ou bloqueiam determinadas informações e imagens, o que basicamente não existe na Internet. Do contrário, o fácil acesso aos meios de comunicação implica “o rápido transbordamento, sufocamento ou atropelamento da informação obtida, assim como a chegada veloz de notícias” (Bauman, 2012, posição 265). As redes sociais se apresentam como espaço propício para a divulgação de mensagens que não foram “supervisionadas”, ampliando-se o escopo de fontes de informação e de desinformação, corroborando de maneira

8 Castells identifica ainda uma terceira característica: a prática de “*spin*”, que corresponde à atividade praticada por políticos, geralmente por meio de assessores, que consiste em comunicar assuntos de forma a favorecer seus interesses ao mesmo tempo em que se infligem danos aos seus oponentes (Castells, 2009).

9 Cf.: “Usuários de *smartphone* devem atualizar WhatsApp, orienta empresa”. Disponível em: <https://agenciabrasil.ebc.com.br/geral/noticia/2019-05/usuarios-de-smartphone-devem-atualizar-whatsapp-orienta-empresa>. Acesso em: 20 ago. 2020.

10 Cf.: “WhatsApp é a principal fonte de informação dos brasileiros, indica pesquisa”. Disponível em: https://www.correiobraziliense.com.br/app/noticia/tecnologia/2019/12/10/interna_tecnologia,812946/whatsapp-e-a-principal-fonte-de-informacao-dos-brasileiros-indica-pes.shtml. Acesso em: 20 ago. 2020.

incisiva para a formação de uma “sociedade da indignação” (Han, 2018a), fomentada pela “política do escândalo” (Castells, 2009).

1.2 FOFOCAS, ESCÂNDALOS E POLÍTICA: DA POLARIZAÇÃO À CRISE DA DEMOCRACIA

Muito antes de nos tornarmos uma sociedade digital, fofocas, rumores e escândalos já eram uma parte crítica do processo de determinação de relações de poder e de mudanças institucionais e sociais (Solove, 2007; Castells, 2009). Como expõe Daniel Solove (2007), desde sempre as pessoas fofocam, espalham rumores e fazem comentários umas sobre as outras. Essas práticas encontraram na Internet um terreno fértil e dele se apossaram, adquirindo novas dimensões. Não sejamos hipócritas: a fofoca faz parte das nossas vidas. E a fofoca não é “inerentemente boa ou ruim – ela possui suas virtudes, assim como seus vícios” (Solove, 2007, p. 11, 12, tradução nossa)¹¹. Nem os escândalos são, se considerarmos o potencial que possuem de nos alertar para situações problemáticas e nas quais, muito frequentemente, nossos direitos estão sendo violados.

A questão é que a recorrência de escândalos corrobora para o advento de uma “sociedade da indignação” (Han, 2018a) no âmago da sociedade da informação. Isso está diretamente ligado à industrialização da atenção humana (Wu, 2017), processo intrínseco às sociedades neoliberais. De acordo com Han,

nos modos atuais de produção imaterial, mais informação e mais comunicação significam mais produtividade, aceleração e crescimento. A informação é uma positividade que, por carecer de interioridade, pode *circular independente do contexto*. Isso permite que a circulação de informações seja acelerada à vontade. (2018b, p. 19)

Busca-se, destarte, acelerar cada vez mais a circulação de informação e comunicação, por meio, por exemplo, de ondas de indignação. Ondas de indignação são um instrumento eficaz em “mobilizar e compactar a atenção” dos cidadãos, mas não são apropriadas para organizar o discurso público, em razão de sua fluidez e volatilidade (Han, 2018a) – “liquidez”, nas palavras de Bauman (2001). Na sociedade de indignação, os escândalos “surtem frequentemente em vista de acontecimentos que têm muito pouca relevância social ou política” (Han, 2018a, posição 174), turvando ainda

11 No original: “Gossip isn't inherently good or evil – it has its virtues as well as its vices”.

mais a esfera pública e inundando o debate político com informações imprecisas, inverídicas e irrelevantes.

Ademais, em se tratando de fofocas, rumores e escândalos, o recente processo de personalização da política traz consequências extraordinárias para as campanhas eleitorais: como as chances de um cidadão optar pelo candidato “A” ou “B” dependem das qualidades que esse candidato possui, campanhas efetivas tendem a enaltecer seus pontos positivos e, concomitantemente, dar destaque às características negativas de seus oponentes (Castells, 2009, p. 238). Uma vez que os aspectos negativos exercem uma maior influência sobre o poder de votos do que aqueles positivos (Castells, 2009, p. 238), a tendência que se observa é o foco cada vez maior nos pontos considerados problemáticos ou polêmicos, fazendo com que a política seja baseada, afinal, em escândalos. Com a Internet, argumenta Solove, a fofoca tem sido reestruturada de maneira a acentuar seus aspectos negativos (Solove, 2007), e acreditamos que o mesmo se aplica aos escândalos.

A comunicação na Internet, especialmente nas redes sociais, contribui para a consolidação da política do escândalo de duas principais maneiras. Primeiramente, a autocomunicação em massa admite alegações e denúncias, verdadeiras ou falsas, de múltiplas fontes, superando a capacidade de “filtragem”¹² que a mídia tradicional possui (Castells, 2009; Han, 2018a). Embora tenhamos acesso a cada vez mais informação, temos sido bombardeados com cada vez mais *má* informação, e “mais informações e mais comunicação não clarificam o mundo” (Han, 2017). Para Solove (2007), temos vivido constantemente em um crepúsculo entre fato e ficção, frequentemente expostos a informações nas quais não podemos (ou não sabemos se podemos) confiar. Mais informações, aponta Han (2017), não geram mais verdade. Pelo contrário: “Quanto mais se liberam informações tanto mais intransparente torna-se o mundo. Por isso, a hiperinformação e a hipercomunicação não trazem luz à escuridão” (Han, 2017, posição 800).

Em segundo lugar, qualquer conteúdo, independentemente do veículo onde seja divulgado, tem o potencial de se tornar viral, de ser rapidamente e descontroladamente difundido. Qualquer pessoa pode postar um informa-

12 Nesse sentido, vale notar que recentemente o governo Bolsonaro realizou uma investida contra a moderação de conteúdo na Internet. Em maio deste ano, o atual presidente da República manifestou a intenção de publicar um decreto regulamentando o Marco Civil da Internet (Lei nº 12.965/2014), no sentido de vetar a moderação de conteúdo por parte das plataformas. Nos termos do decreto, qualquer conteúdo – por exemplo, discursos antivacina, desinformação eleitoral, *deepfakes* – só poderá ser removido das redes sociais após decisão judicial, única e exclusivamente, “invertendo a lógica da liberdade” (Souza, 2021).

ção na Internet livremente (Solove, 2007, p. 189). Em um mundo onde se torna cada vez difícil separar o que é verdadeiro do que é falso, rumores e escândalos podem se espalhar rapidamente, de modo que a Internet acaba sendo empregada como uma “ferramenta poderosa para perpetrar ataques maliciosos contra pessoas e ideias” (Solove, 2007, p. 36).

Tudo isso ocorre em um contexto no qual cada clique pode ser monetizado e no qual, consequentemente, maior engajamento significa maior lucro. E os maiores engajamentos, sabemos, surgem a partir da divulgação de notícias negativas e/ou, muitas vezes, falsas (Citron; Chesney, 2019). Seguindo a lógica capitalista subjacente à política midiática, estabelece-se um verdadeiro mercado de conteúdos prejudiciais, o qual, na ausência de materiais autênticos aptos a gerar escândalos, é suprido com insinuações ou até mesmo fabricação de notícias (Castells, 2009, p. 249). Nas palavras do autor,

trata-se de inserir negatividade de conteúdos na imagem da pessoa que se quer destruir, a fim de eliminar o vínculo de confiança com os cidadãos. Daí a prática de operadores políticos profissionais no sentido de buscar materiais prejudiciais para determinados líderes políticos, manipulando-os e até fabricando-os para aumentar o efeito destrutivo. (Castells, 2018b, localização 260)

Evidências empíricas sugerem que escândalos políticos influenciam o comportamento dos eleitores de maneira distinta, variando em razão de fatores como o país ou o cargo em disputa (Castells, 2009, p. 250). Sucintamente, é possível afirmar que

os efeitos da política do escândalo sobre resultados políticos específicos são em grande parte indeterminados. Eles dependem do contexto cultural e institucional, da relação entre o tipo de escândalo e o político envolvido no escândalo, do clima social e político do país e da intensidade do efeito do cansaço detectado entre os cidadãos após a repetição interminável de histórias de escândalo na mídia. (Castells, 2009, p. 253, tradução nossa)¹³

13 No original: “The effects of scandal politics on specific political outcomes are largely undetermined. They depend on the cultural and institutional context, the relationship between the kind of scandal and the politician involved in the scandal, the social and political climate of the country, and the intensity of the fatigue effect detected among citizens after endless reiteration of scandal stories in the media. Effects must also be measured over time, and are often indirect in their manifestations; for example, another politician suffering consequences by association”.

Uma coisa, porém, é certa. Independentemente dos resultados específicos em distintos contextos, a política do escândalo promove a transformação geral do cenário político ao fazer com que os cidadãos passem a associar automaticamente a política aos comportamentos escandalosos exibidos pela mídia e compartilhados nas redes sociais, o que, por sua vez, gera um contexto de insatisfação e desconfiança para com instituições e classes políticas, acentuando a crise de legitimidade que hoje se verifica por todo o mundo (Castells, 2009, p. 254). Tal crise possui um aspecto ineludivelmente partidário (Benkler; Faris; Hal, 2018), como explicaremos brevemente a seguir.

1.2.1 Ciberpolarização

Na sociedade da informação, aponta Han, ocorre um aplainamento da comunicação: uma vez que o “segredo, o estranhamento ou a alteridade representam barreiras à comunicação ilimitada” (2018b, p. 20) são elementos a ser eliminados em nome da aceleração do fluxo de informações. A negação à alteridade ou ao estranhamento, porém, não gera uma conformidade de opiniões, mas se converte em uma “positividade da diferença ou da diversidade” (Han, 2018b), corroborando a polarização. A integração e a divisão são, segundo Bauman (2012), processos mutuamente complementares, e a fragmentação política não é um impedimento à circulação da informação.

Não obstante, Cass Sunstein argumenta que, em uma democracia funcional, as pessoas não devem viver em “câmaras de eco” ou “casulos informacionais”; do contrário, elas devem ser expostas a assuntos, ideias e pontos de vistas distintos, ainda que conscientemente não desejem essa exposição. Embora, em diversos contextos, o ser humano naturalmente tenha uma tendência à homofilia, a se conectar e a criar laços com pessoas com as quais se identificam (Sunstein, 2018, posição 83), a Internet e as redes sociais têm influenciado de maneira negativa esse viés. Para Sunstein,

novas tecnologias, enfaticamente incluindo a Internet, tornam mais fácil para as pessoas se cercar (virtualmente, é claro) das opiniões de outras que pensam de maneira semelhante, mas que estariam, do contrário, isoladas, e se isolar de opiniões conflitantes. Por essa razão somente, elas são um terreno fértil para a polarização, e um terreno fértil tanto para a democracia quanto para a paz social. (Sunstein, 2007, p. 63, 64, tradução nossa)¹⁴

14 No original: “New technologies, emphatically including the Internet, make it easier for people to surround themselves (virtually of course) with the opinions of like-minded but otherwise isolated others, and to

A construção de comunidades homogêneas é um impulso voluntário facilitado pelos ambientes virtuais, mas que tende a ser negativamente exacerbado com as redes sociais e as bolhas do filtro, “favorecendo o isolamento e o entrincheiramento de grupos sociais e políticos” (Magrani, 2014, p. 126). Conquanto o autoisolamento e a personalização de conteúdos não sejam fenômenos intrinsecamente negativos, eles corroboram a disseminação de rumores e conteúdos falsos, promovendo a fragmentação e polarização da sociedade (Sunstein, 2018, posição 192). Para Bauman, as novas tecnologias de informação e comunicação, ao anular as distâncias temporais e espaciais, não homogeneizam a condição humana, mas sim tendem a polarizá-la, fragmentando-a cada vez mais (Bauman, 2012, posição 298).

É importante deixar claro que a Internet não é causadora da polarização (nem das fofocas, da política do escândalo, etc., como buscamos demonstrar nos tópicos anteriores). Yochai Benkler, Robert Faris e Hal Roberts argumentam que o “Facebook não criou a arquitetura assimétrica da esfera pública americana” (2018, posição 5586, tradução nossa), e que nem a Internet ou a blogosfera o fizeram, pois essa assimetria sempre esteve clara em se tratando de redes de televisão e rádio. Apesar de os autores discorrerem acerca do contexto estadunidense, e especificamente sobre o Facebook, é possível vislumbrar como essas considerações se aplicam a outras democracias e redes sociais.

Ainda assim, a “polarização de grupos está inquestionavelmente ocorrendo *online*” e a Internet, para muitas pessoas, está servindo como um terreno fértil para o extremismo (Sunstein, 2018, posição 1323). A polarização de grupos se refere, segundo Sunstein, a algo simples, que é o fato de as pessoas, mesmo após deliberações e a exposição a visões opostas, tenderem a adotar uma posição ainda mais extrema no sentido daquilo que os membros de determinado grupo originalmente acreditavam ou tomavam como certo (Sunstein, 2007, p. 59-60; 2018, posição 1183)¹⁵. Nessa perspectiva, e no contexto da esfera pública conectada, é possível falar na existência de uma “ciberpolarização” (Sunstein, 2018).

insulate themselves from competing view. For this reason alone, they are a breeding ground for polarization, and potentially dangerous for both democracy and social peace”.

- 15 Nas palavras de Sunstein, “[...] the term ‘group polarization’ refers to something very simple: after deliberation, people are likely to move toward a more extreme point in the direction to which the groups members were originally inclined. With respect to the internet and new communications technologies, the implication is that groups of like-minded people, engaged in discussion with one another, will end up thinking the same thing that they thought before – but in more extreme form” (2007, p. 59-60).

Embora a comunicação em massa atinja sua velocidade máxima quando pessoas com pensamentos e ideais semelhantes trocam informações, em uma verdadeira “reação em cadeia do igual” (Han, 2017), não podemos creditar a polarização somente à Internet e às redes sociais (ou à mídia tradicional). Por mais que sejam atores que influenciam a polarização, principalmente por meio da criação e manutenção de filtros-bolha¹⁶, boa parte da responsabilidade pela fragmentação social está em nossas próprias mãos. Sem dúvidas, a Internet tende a reduzir a amplitude das discussões, mas isso se dá também por culpa dos próprios usuários (Magrani, 2014, p. 126). Nas palavras de Solove, “quando se trata de fofocas e rumores na internet, os culpados somos nós mesmos” (2007, tradução nossa)¹⁷.

1.2.2. Crise da democracia

A predominância da política midiática e da política do escândalo na condução de assuntos públicos traz como resultados uma maior exposição da corrupção das instituições políticas e o consequente declínio da confiança dos cidadãos nessas instituições (Castells, 2009, p. 289), *contribuindo* para uma crise global da legitimidade política. Crise essa que resulta do rompimento do vínculo subjetivo entre os desejos e as ideias dos cidadãos e as ações dos representantes eleitos, sendo “precisamente o crescimento da distância entre as crenças nas instituições políticas e o desejo por ação política [o] que constitui a crise da democracia” (Castells, 2009, p. 295, tradução nossa)¹⁸. Em uma perspectiva mais ampla, a crise da legitimidade e as suas consequências para a democracia estão relacionadas à crise da Nação-Estado na sociedade global interconectada, uma das decorrências dos processos de globalização e identificação, que são contraditórios entre si (Castells, 2009, p. 296). A crise de representação se une, afinal, à crise identitária resultante da globalização (Castells, 2018b, localização 171). À medida que cresce a distância entre governantes e cidadãos, diminui a capacidade que os governos possuem de conciliar seus objetivos globais e as demandas locais ou nacionais relativas à aplicação de recursos e à própria legitimidade política.

16 Retornaremos ao assunto no tópico 2.2.

17 No original: “When it comes to gossip and rumor on the Internet, however, the culprit is ourselves”.

18 No original: “It is precisely this growing distance between belief in political institutions and desire for political action that constitutes the crisis of democracy”.

O senso comum (mas também muitos acadêmicos, cientistas políticos, etc.) enxerga o declínio da confiança dos cidadãos nas instituições democráticas como uma consequência da Internet e, mais precisamente, do advento das redes sociais. Contudo, interpretar que a crise democrática é uma *consequência* dos novos meios de comunicação é uma conclusão precipitada e, acima de tudo, equivocada. Trata-se de uma crise institucional, não meramente tecnológica (Benkler; Farris; Hal, 2018). De fato, diversos autores apontam para a existência de um declínio contínuo e de longa data da confiança dos cidadãos nas instituições democráticas e elites políticas (Hofmann, 2019). E hoje, em razão da mediatização política, o que ocorre é a atribuição de um novo papel para a esfera pública (conectada), concomitantemente a uma “mudança fundamental das práticas democráticas”:

A mudança geral da confiança para a desconfiança transformou a esfera pública em um espaço de observação, avaliação, controle e de escândalo de atores políticos. Entretanto, o ponto a ser destacado no contexto da noção de democracia midiática é que a mudança de práticas e atitudes democráticas pode ser interpretada como um motor da evolução das redes sociais e, em um sentido mais amplo, da mediatização. (Hofmann, 2019, p. 10, tradução nossa)¹⁹

As tecnologias de informação e comunicação constituem um espaço de possibilidades que deriva sua forma específica, principalmente da contínua transformação dos próprios processos democráticos, de modo que os serviços de comunicação digital servem para a experimentação de novos modos de expressão política, ao mesmo tempo em que são moldados por eles (Hofmann, 2019). Neste contexto, parece-nos imprescindível discutir o papel da política midiática, uma vez que ela, frequentemente, busca “inserir negatividade de conteúdos na imagem da pessoa que se quer destruir, a fim de eliminar o vínculo de confiança com os cidadãos”, inclusive com a manipulação ou fabricação de materiais prejudiciais, a fim de aumentar seu efeito (Castells, 2018b, localização 260).

A dissociação entre as práticas dos governantes eleitos e as expectativas de seus jurisdicionados é, então, exacerbada pela política midiática e pela política do escândalo, gerando não só maior descontentamento entre

19 No original: “The general shift from trust to distrust has turned the public sphere into a space of watching, evaluating, controlling and scandalising political actors and actions. However, the point to be highlighted in the context of the notion of mediated democracy is that the changing democratic practices and attitudes can be interpreted as one driver of the evolution of social networks and, in a wider sense, of mediatisation”.

os cidadãos, como também desafios cada vez maiores à própria ordem democrática. Nesse sentido, a política midiática e a política do escândalo, acentuadas pela polarização da esfera pública conectada, estão diretamente ligadas à crise – “ruptura”, nas palavras de Castells (2018b) – da democracia, e possivelmente as *deepfakes* irão acelerar ainda mais esse processo.

2 DEEPFAKES: CATALISADORES DA RUPTURA?

A Era da Informação é marcada pela interconexão da comunicação humana e pela disponibilização, em rede, de basicamente toda e qualquer informação, o que cria um “universo midiático no qual estamos permanentemente imersos” (Castells, 2018b, localização 246). Nós construímos nossa noção de realidade a partir do intercâmbio de informações realizado nesse universo, o que, por conseguinte, molda nosso comportamento e nossas decisões (Castells, 2018b). Como procuramos demonstrar no tópico anterior, fofocas e rumores fazem parte dessa construção: às vezes de maneira benéfica, às vezes de maneira prejudicial. Mas ainda que estudiosos apontem que a fofoca é essencial para o controle social, as pessoas (incluindo-se aqui os políticos) frequentemente fofocam de maneira não benéfica para a sociedade. Do contrário, o fazem em nome de seu próprio interesse e em detrimento do bem comum (Solove, 2007, p. 65).

Em se tratando da comunicação realizada por candidatos e instituições políticas em disputas eleitorais, a troca de informações ocorre, como exposto nos tópicos anteriores, tanto na mídia tradicional quanto nos meios de autocomunicação em massa, sobretudo nas redes sociais. Independentemente do meio onde se veicule determinada informação, fato é que, para se atingir um grande número de pessoas, é necessária uma mensagem simples e impactante. A mensagem mais simples que pode haver é uma imagem; por sua vez, a imagem mais simples, e com a qual a maioria das pessoas se identifica, é um rosto humano (Castells, 2009, p. 204). Essa identificação, por sua vez, resulta na construção de uma relação de confiança (Castells, 2018b, localização 251), essencial para a efetividade da comunicação. Nesse sentido, importante considerar que a efetividade de uma mensagem audiovisual no que se refere à sua capacidade de convencimento está diretamente relacionada à maneira com a qual nosso cérebro processa informações (Castells, 2009, p. 196). E entre os diversos fatores que influenciam nosso processo de convencimento e tomada de decisões encontra-se a confiança.

Nas sociedades democráticas atuais, a luta pelo poder passa pela política midiática e pela política do escândalo (Castells, 2018b, localização 245), que se valem diretamente da imagem dos candidatos, mas sobretudo de seus oponentes. Na ausência de escândalos “verdadeiros”, muitas das vezes as campanhas políticas apelam para o uso de informações imprecisas, fora de contexto ou evidentemente falsas, as chamadas *fake news* – prática que se tornou corriqueira a partir de meados da década passada em todo o mundo²⁰. Já no início do século, Solove (2007, p. 35, 36) alertava para o risco do compartilhamento de conteúdo digitalmente adulterado e das proporções que esse compartilhamento poderia alcançar com a evolução da Internet. A questão é que hoje as campanhas de desinformação se elevaram a um novo patamar, e as *deepfakes* podem tornar o jogo político ainda mais complicado.

2.1 AFINAL, O QUE SÃO DEEPFAKES?

A sigloneização dos termos “*deep learning*”²¹ (aprendizado profundo, um ramo do aprendizado de máquina) e “*fake*” (falso) deu origem à expressão “*deepfake*”, que costuma ser empregada heurísticamente, referindo-se à ampla gama de falsificações hiper-realistas de imagens, vídeos e áudios realizadas digitalmente. Ao utilizar algoritmos de *machine learning*²² para inserir rostos e vozes em gravações de vídeo e áudios reais, tecnologias de *deepfakes* possibilitam a criação de personificações extremamente realistas, cujo resultado final é uma mídia dotada de tamanha verossimilhança, a ponto de se tornar (quase) impossível distinguir o que é real do que aquilo

20 Neste trabalho, não discutiremos com profundidade a questão da regulação das *fake news*. Vale mencionar, contudo, que o debate em torno do assunto no Brasil está concentrado principalmente em torno do Projeto de Lei nº 2.630/2020, popularmente conhecido como “PL das *Fake News*”. Duramente criticado por especialistas e instituições da sociedade civil, o texto foi aprovado pelo Senado em junho de 2020, e está aguardando a análise da Câmara dos Deputados e da sanção da Presidência da República para entrar em vigor.

21 Trata-se de um ramo do aprendizado de máquina (conceituado na nota abaixo), isto é, “corresponde a uma técnica de inteligência artificial que realiza o treinamento das chamadas ‘redes neurais profundas’ (*deep neural networks*, ou DNN, em inglês), algoritmos que se comportam de maneira semelhante aos neurônios em nosso cérebro. As redes neurais profundas são formadas por um grande conjunto de neurônios artificiais, que, assim como os neurônios que possuímos, realizam operações simples, se considerados individualmente, mas, em conjunto, são capazes de realizar atividades complexas e não-lineares, como reconhecer o rosto de uma pessoa a partir de uma imagem projetada em uma tela” (Oliveira, 2021, p. 198).

22 A expressão “*machine learning*” pode ser traduzida como “aprendizado de máquina”, e se refere, segundo Mulholland e Frajhof, a “sistemas de Inteligência Artificial (IA) que desenvolvem a capacidade de tomadas de decisão absolutamente autônomas em relação à interferência humana” (2019, p. 265, 266). Em outras palavras, “torna-se possível por meio de tratamento de dados em massa – *inputs* – o desenvolvimento de autoaprendizagem das máquinas – *i.e.* programas e sistemas – que permite o alcance de determinados resultados – *outputs* –, independentemente de qualquer mediação por um ser humano” (Mulholland; Frajhof, 2019, p. 266).

que foi digitalmente produzido. Colocando em termos mais simples, uma *deepfake* é um tipo de mídia – incluindo-se imagens, áudio e vídeo – que é ou manipulada ou inteiramente gerada por uma inteligência artificial (Schick, 2020, p. 8).

Nesse ponto, destaca-se a distinção realizada por Nina Schick: “Quando utilizada maliciosamente como desinformação, ou quando utilizada como má informação, uma mídia sintética é chamada *deepfake*” (Schick, 2020, p. 8, tradução nossa)²³. Outra diferenciação cabível é aquela que se estabelece entre *deepfakes*, no sentido trazido *supra*, e as chamadas *cheapfakes*²⁴. Pensemos, por exemplo, na imagem divulgada da presidente Dilma Rousseff supostamente comendo um pombo vivo, na cidade Osasco²⁵. Ou, então, nas imagens que circularam da candidata à vice-presidência nas eleições de 2018, Manuela D’Ávila, em que apareciam suas supostas tatuagens dos rostos de Che Guevara e Lênin e que, ainda, mostravam a política utilizando uma faixa presidencial claramente editada²⁶. Ainda que sejam manipulações grotescas, ou, no termo exposto antes, *cheapfakes*, as imagens geraram certo nível de “credibilidade”, tendo sido tão compartilhadas por usuários de redes sociais que diversos *sites* de verificação de conteúdo tiveram que esclarecer que se tratavam de montagens. Desconsideradas as situações em que as manipulações realizadas digitalmente são facilmente perceptíveis, ou até mesmo esdrúxulas, cabe-nos agora analisar quais são, de fato, as características das *deepfakes*, a fim de melhor entender por que essa inovação tecnológica tem suscitado tantos debates e preocupações.

A despeito de diferentes classificações empregadas, a maioria dos autores aponta dois aspectos fundamentais das *deepfakes*: a *qualidade* das mídias produzidas e a *acessibilidade* às tecnologias de manipulação digital. Em relação ao primeiro aspecto, Schick afirma que a “IA vai criar efeitos audiovisuais muito melhores do que qualquer coisa que os estúdios de CGI²⁷ já fizeram no passado” (Schick, 2020, p. 49). Talvez o mais conhecido

23 No original: “When used maliciously as disinformation, or when used as misinformation, a piece of synthetic media is called a ‘deepfake’”.

24 O termo “*deepfake*” pode ser traduzido como “falsificação profunda”, ao passo que “*cheapfake*” pode ser entendido como “falsificação barata”.

25 Cf.: “Mentira que Dilma comeu pombo em calçada de Osasco”. Disponível em: <https://boainformacao.com.br/2019/07/mentira-que-dilma-comeu-pombo-em-calçada-de-osasco/>.

26 Cf.: “Imagem de Manuela D’Ávila com tatuagens de Che Guevara e Lênin é montagem”. Disponível em: <https://www.otempo.com.br/hotsites/elei%C3%A7%C3%B5es-2018/imagem-de-manuela-d-avila-com-tatuagens-de-che-guevara-e-lenin-e-montagem-1.2033630>.

27 “CGI” é a sigla para “*computer-generated imagery*”, expressão que pode ser traduzida como “imagens geradas por computador”.

exemplo do uso de GCI seja o filme “O curioso caso de Benjamin Button”, de 2008, que custou cerca de 170 milhões de dólares à época²⁸. Hoje, pouco mais de dez anos após o lançamento do filme, o processo de manipulação de imagens tornou-se mais eficiente, devido, sobretudo, aos avanços proporcionados pela IA. O trabalho que antes demandava a atuação de diversos especialistas em computação gráfica agora pode ser realizado mediante o uso de *softwares* de computador e até mesmo *apps* para aparelhos celulares.

A segunda característica marcante das *deepfakes* é a acessibilidade (Kietzmann et al., 2020). Há até relativamente pouquíssimo tempo esse tipo de tecnologia de manipulação digital de imagens era acessível apenas para grandes estúdios de Hollywood, como expusemos antes. Hoje, todavia, qualquer pessoa pode baixar um dos diversos *apps* disponíveis no mercado e, com poucos passos, produzir uma *deepfake* (Maras; Alexandrou, 2019). É possível falar, portanto, na democratização do acesso (Schick, 2020, p. 49) a essas tecnologias: à medida que a tecnologia melhora e se torna mais acessível por meio de aplicativos e *softwares*, mais pessoas poderão acessá-la – o que, conseqüentemente, trará aspectos positivos e negativos.

Em suma, o ponto de inflexão dessas tecnologias é o escopo, a escala e a sofisticação dos métodos e instrumentos envolvidos, os quais permitem que basicamente qualquer pessoa com acesso a um computador produza conteúdos artificiais praticamente indistinguíveis daqueles considerados autênticos (Westerlund, 2020). Combinados, esses fatores são uma “receita para o sucesso” das *deepfakes* (Kietzmann et al., 2020).

2.2 DEEPFAKES, EROSÃO DA CONFIABILIDADE E O “DIVIDENDO DO MENTIROSO”

Distinguir vídeos autênticos daqueles artificialmente criados ou modificados tem se tornado cada vez mais difícil, o que se dá, obviamente, em razão de sua crescente qualidade (Kietzmann et al., 2020). Nessa perspectiva, importante ressaltar que a verossimilhança das *deepfakes* está diretamente relacionada ao aumento de sua credibilidade. Mídias sintéticas estão se tornando mais críveis e o seu impacto mais significativo. Isso se dá porque, embora nossa confiança em mídias fotográficas tenha decaído durante

28 Cf.: “What’s Wrong With ‘Benjamin Button?’”. Disponível em: https://www.vulture.com/2008/11/whats_wrong_with_benjamin_but.html. Acesso em: 20 ago. 2020.

às últimas décadas, nós ainda depositamos muita confiança em evidências videográficas (Kietzmann et al., 2020).

Alguns dos potenciais aspectos negativos das tecnologias de criação de mídias sintéticas são, de certa forma, evidentes. É o caso das hipóteses mencionadas na introdução: fraudes, furto de dados biométricos e pornografia de vingança. Todavia, há também questões subjacentes, mas nem por isso menos importantes. Dedicaremos nossa atenção neste tópico a duas delas, nomeadamente a erosão da confiabilidade e o “dividendo do mentiroso” (Citron; Chesney, 2019).

Luciano Floridi (2018) afirma que “as tecnologias digitais parecem minar nossa confiança na natureza original, genuína e autêntica daquilo que vemos e ouvimos”. Devido à facilidade de se obter e difundir informações falsas através, principalmente, de redes sociais, tem se tornado progressivamente mais difícil saber no que confiar, o que traz consequências prejudiciais no que tange à tomada de decisão informada e ao estabelecimento de um diálogo público esclarecido. Teorias conspiratórias, rumores e desinformações tendem a se espalhar rapidamente (Westerlund, 2020), ao mesmo tempo em que o corrente “apocalipse informacional”²⁹ faz com que as pessoas pensem que não podem confiar em informações que não sejam provenientes de suas redes de contato, nas quais se incluem seus familiares, amigos próximos ou colegas de trabalho, por exemplo.

Na esteira do que apontam Danielle Citron e Robert Chesney (2019), podemos vislumbrar três fenômenos que ajudam a compreender por que as *deepfakes* possuem uma inclinação para se tornarem virais. O primeiro deles diz respeito ao fato de que as pessoas, por vezes, não dedicam atenção suficiente à obtenção de informação, confiando naquilo que dizem outras pessoas que já teriam se informado sobre determinado assunto e tomando isso como suficiente. Uma vez que as pessoas não conseguem se informar de maneira adequada sobre tudo, elas frequentemente acreditam no que os outros disseram, presumindo que estes hajam analisado adequadamente as informações, ainda que a nova informação recebida vá de encontro àquilo que já sabiam. Trata-se, nas palavras de Cass Sunstein, da “cascata informacional” (2007, p. 84)³⁰.

29 Schick utiliza o neologismo “*infocalypse*”, o qual define como “*the increasingly dangerous and untrustworthy information ecosystem within which most humans now live*” (2020, p. 9).

30 “*In an informational cascade, people cease relying, at a certain point, on their private information or opinions. They decide instead on the basis of the signals conveyed by others. It follows that the behavior*

O segundo, por sua vez, se refere a nossa tendência natural de atração por informações novas e negativas, o que faz com que boatos ou rumores – que, geralmente, são negativos – sejam muito mais compartilhados do que outros tipos de informação (Meyer, 2018). Alguns pesquisadores, como Marc Trusler e Stuart Soroka (2014), se referem a esse fenômeno como “*negativity bias*”. A partir de um estudo empírico, os autores constataram que pessoas interessadas por política possuem certa inclinação a optar por ler notícias negativas, ainda que conscientemente declarem o contrário³¹. Em outro estudo, que analisou o compartilhamento de notícias no Twitter entre 2006 e 2017, concluiu-se que boatos e rumores falsos alcançavam as pessoas em uma velocidade até dez vezes maior que notícias verdadeiras; curiosamente, concluiu-se também que, ao contrário do que imagina o senso comum, *bots* espalham notícias verdadeiras e falsas com na mesma proporção, o que indica que os principais responsáveis pela divulgação de *fake news* são, de fato, os seres humanos (Vosoughi; Roy; Aral, 2018).

Finalmente, a “bolha dos filtros”³² (Pariser, 2012) agrava a disseminação de notícias falsas ou ruins, pois os algoritmos das plataformas de comunicação tendem a enfatizar conteúdos “populares”, especialmente se houverem sido compartilhados por amigos ou conexões realizadas nas redes, fazendo com que informações relativamente homogêneas permeiem nossos *feeds*. Isso porque, para poder lidar com o excesso de informação ao qual é submetido atualmente, o leitor “médio” passa a depender da filtragem realizada pelos algoritmos (Hildebrandt; De Vries, 2013), o que frequentemente vem associado à criação de “câmaras de eco” (Sunstein, 2007) ou “bolhas do filtro” (Pariser, 2012).

Esses três fatores, somados ao fato de que as *deepfakes* têm se tornando tão realísticas a ponto de confundir um espectador leigo ou desatento, podem levar, nas palavras de Marc Blitz (2018), a uma “anarquia informacional”. Isto é, a um mundo caracterizado por uma insuperável paranoia acerca do que é verdadeiro e do que é falso, o que afeta profundamente os processos de tomada de decisão individuais e coletivos que dependem da

of the first few people can, in theory, produce similar behavior from countless followers. [...] Even among specialists and indeed doctors, cascades are common.” (Sunstein, 2007, p. 84, 85)

31 “Our results suggest that, regardless of their preferences as stated in a survey, participants are more likely to select negative content. The bias toward negative content is greater for politically interested respondents, and politically interested participants are drawn to strategic frames as well.” (Trusler; Soroka, 2014, p. 373)

32 Eli Pariser (2011) utiliza o termo “*filter bubble*”, que é traduzido para o português como “bolha dos filtros” (Pariser, 2012) e, frequentemente, “filtros-bolha”.

existência de um mínimo de confiança em fontes externas de comunicação (Blitz, 2018). *Deepfakes* podem causar tamanho desgaste à credibilidade das instituições, tanto públicas quanto privadas, que essas possivelmente enfrentarão uma crescente dificuldade de se mostrarem confiáveis para as pessoas de maneira geral, caracterizando a erosão da confiabilidade. Nas palavras de Castells,

embora os efeitos da política do escândalo sobre políticos específicos sejam indeterminados, tal política gera um efeito secundário que é devastador: o de inspirar o sentimento de desconfiança e reprovação moral sobre o conjunto dos políticos e da política, contribuindo assim para a crise de legitimidade. (2018b, localização 268)

É inegável, portanto, a possibilidade de *deepfakes* afetarem a confiança que os cidadãos depositam nas autoridades públicas, uma vez que conteúdos forjados poderiam mostrar políticos ou agentes com funções administrativas dizendo coisas absurdas, praticando atos imorais, etc. Isso é ainda mais inquietante quando consideramos o papel de chefes de Estado e instituições governamentais na produção e difusão de conteúdo falso³³. Sobretudo se o governo se constituir em um regime despótico ou autoritário, há um verdadeiro incentivo para que se subvertam os meios de informação populares, de modo que apenas as pessoas no poder tenham a capacidade de controlar narrativas (Schick, 2020, p. 125).

Outra questão a ser considerada é que enquanto compartilhar uma informação é extremamente fácil, corrigir uma informação incorreta que foi compartilhada é muito mais difícil (Westerlund, 2020), lógica que se aplica às *deepfakes*. Na maioria das vezes, o alcance da informação verídica ou ajustada não é o mesmo que obteve o boato ou a desinformação; em muitas outras, a correção pode ser inapta para sanar o dano causado inicialmente (Citron; Chesney, 2019). Basta imaginar a hipótese em que uma *deepfake* retratando determinado candidato seja divulgada às vésperas da eleição à qual está concorrendo. A existência de um curto lapso temporal entre a divulgação da notícia e o comparecimento dos eleitores às urnas poderia, inclusive, “impossibilitar o candidato ofendido de esclarecer os fatos a seus

33 Em três anos de governo, de 2017 a janeiro de 2020, Donald Trump, presidente dos EUA, havia feito mais de 18 mil alegações falsas ou enganosas (Shick, op. cit., p. 93). No Brasil, Jair Bolsonaro atingiu a marcar de mil declarações distorcidas ou falsas com 492 dias de governo (Cf.: “Bolsonaro chega a 1.000 declarações falsas ou distorcidas com 492 dias de mandato”. Disponível em: <https://www.aosfatos.org/noticias/bolsonaro-chega-1000-declaracoes-falsas-ou-distorcidas-com-492-dias-de-mandato/>. Acesso em: 20 ago. 2020).

eleitores ou de conseguir esclarecer o fato, por não haver tempo hábil de que o vídeo resposta se propague e tenha a mesma escalabilidade do vídeo falso” (Rais; Sales, 2020, localização 499). Bauman (2012) nos aponta que o movimento da informação ganha velocidade em um ritmo muito mais acelerado que a mudança da situação sobre a qual se informa. Em outras palavras, “a falsidade voa, e a verdade vem mancando atrás dela” (Swift, 1710).

Não devemos nos esquecer também que diversas pessoas provavelmente irão tomar por fraudulentos inúmeros vídeos genuínos, simplesmente porque assumirão a postura de afirmar que “qualquer coisa na qual não queiram acreditar deve ser *fake*” (Westerlund, 2020). É o que Chesney e Citron denominam “dividendo do mentiroso”³⁴: as pessoas estarão imbuídas em tamanho ceticismo que duvidarão da autenticidade de quaisquer evidências em áudio ou vídeo, mesmo as reais. Esse ceticismo poderá ser utilizado para desacreditar tanto conteúdos autênticos quanto conteúdos de fato adulterados ou imprecisos, como frequentemente feito por políticos como Donald Trump, nos EUA, e Jair Bolsonaro, no Brasil.

Diante do exposto, podemos considerar que o maior problema ocasionado pelas *deepfakes* não é a desinformação *per se*, mas sim a sensação de desconfiança e insegurança causada pelo constante contato com informações falsas, o que fará que nem mesmo conteúdos audiovisuais sejam tidos como fontes confiáveis de informação. Essa é a concretização do que Schick definiu como o “apocalipse informacional”. Nossa maior ameaça não se encontra na possibilidade de sermos, eventualmente, enganados, senão no fato de que tenderemos a considerar todas as informações que nos forem apresentadas como sendo errôneas, fictícias ou inverídicas.

CONSIDERAÇÕES FINAIS: NEM TUDO ESTÁ PERDIDO

A Internet e a redes sociais não causaram o declínio da democracia representativa, nem dos meios de expressão e participação política. E dificilmente as *deepfakes* o farão. Na esteira do que argumenta Jeanette Hofmann, devemos entender a Internet como um campo de treinamento para se experimentar novas formas de agir democrático, tendo em mente que muitas das vezes os experimentos que realizamos (como as *deepfakes*) não deixarão as democracias inalteradas (Hofmann, 2019, p. 13). Não se trata, portanto, de

34 Chesney e Citron (2019) utilizam o termo “*liar’s dividend*”.

tentar prevenir ou abolir todos os efeitos negativos das *deepfakes*, mas sim de pensar como controlá-los e mediar seus efeitos problemáticos³⁵.

Se, para Castells, a política do escândalo é a epítome da política midiática (2009, p. 8), não parece exagero afirmar que as *deepfakes* poderão se tornar, em breve, a epítome da política do escândalo, aprofundando ainda mais a crise de confiança e a crise de legitimidade que assolam a maioria das sociedades democráticas. A era da política midiática e da política do escândalo, da pós-verdade e das *fake news*, ao que tudo indica, está longe de acabar. Do contrário, com o surgimento das ditas *deepfakes*, a desinformação parece adquirir ímpeto, colocando-se como uma questão cada vez mais desafiadora.

Todavia, devemos levar em consideração que as *deepfakes* não fazem muito além de aprofundar problemas já existentes em nossa sociedade. Não é de hoje que as instituições políticas vêm sofrendo com a crise de confiança, ocasionada, sobretudo, pela prática reiterada e ininterrupta da política do escândalo; tampouco é novidade que, em razão dessa crise de confiança, acentuou-se a crise de legitimidade da própria democracia representativa, bem como o fortalecimento do processo de ruptura democrática. Nessa perspectiva, para combater as *deepfakes*, primeiro é preciso reestruturar nossas instituições e reconstruir a confiança dos cidadãos – o que, não nos enganemos, não é tarefa fácil. As *deepfakes* simbolizam um ponto de inflexão no âmbito da prática democrática tal qual a conhecemos, o que não é necessariamente ruim. Ainda há tempo para ação antes que a democracia venha, de fato, a ruir. E talvez as *deepfakes* estejam aí para nos lembrar disso.

Muitos são os desafios que se colocam a nossa frente. Serão necessárias discussões aprofundadas no que se refere aos limites da liberdade de expressão, à regulação tecnológica e da atividade econômica, à educação midiática. Reconhecendo as limitações da presente pesquisa, que não pôde se debruçar sobre as eventuais soluções para os problemas apresentados, este artigo pretendeu ser um pontapé inicial em um debate tão caro quanto urgente.

35 É a sugestão feita por Solove (2007, p. 65) ao discorrer sobre a possibilidade de se regular a disseminação de fofocas e rumores na Internet, mas que acreditamos ser plenamente cabível no caso das *deepfakes*.

REFERÊNCIAS

BAUMAN, Zygmunt. *Modernidade líquida*. Trad. Plínio Dentzien. Rio de Janeiro: Zahar, 2011. Edição Digital.

_____. *Globalização: as consequências humanas*. Trad. Marcus Penchel. Rio de Janeiro: Zahar, 2012. Edição Digital.

BENKLER, Yochai. *The wealth of networks: how social production transforms markets and freedom*. New Haven [Conn.]: Yale University Press, 2006.

_____; FARIS, Robert; ROBERTS, Hal. *Network propaganda: manipulation, disinformation, and radicalization in American politics*. New York, NY: Oxford University Press, 2018.

BLITZ, Marc Jonathan. Lies, Line Drawing, and (Deep) Fake News. *Oklahoma Law Review*, 2018. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3328273. Acesso em: 20 ago. 2020.

CASTELLS, Manuel. *O poder da identidade*. Trad. Klauss Brandini Gerhardt. São Paulo: Paz e Terra, 2018a.

_____. *A galáxia da internet: reflexões sobre a internet, os negócios e a sociedade*. Trad. Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2003.

_____. *Communication Power*. Nova York: Oxford University Press, 2009.

_____. *Redes de indignação e esperança: movimentos sociais na era da internet*. Trad. Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013.

_____. *Ruptura: a crise da democracia liberal*. Trad. Joana Angélica d'Ávila Melo. Rio de Janeiro: Zahar, 2018b.

CITRON, Danielle; CHESNEY, Bobby. Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, v. 107, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954. Acesso em: 20 ago. 2020.

FLORIDI, Luciano. Deepfakes and a Future of Ectypes. *Philosophy & Technology*, v. 31, p. 317-321, 2018. Disponível em: <https://doi.org/10.1007/s13347-018-0325-3>. Acesso em: 20 ago. 2020.

HAN, Byung-Chul. *Sociedade da transparência*. Trad. Enio Paulo Giachini. Petrópolis: Vozes, 2017.

_____. *No enxame: perspectivas do digital*. Trad. Lucas Machado. Petrópolis/RJ: Vozes, 2018a.

_____. *Psicopolítica: o neoliberalismo e as novas técnicas de poder*. Trad. Maurício Liesen. Belo Horizonte: Âyiné, 2018b.

HAO, Karen. An AI app that “undressed” women shows how deepfakes harm the most vulnerable. *MIT Technology Review*, [s.l.], 28 jun. 2019a. Disponível em: https://www.technologyreview.com/s/613898/an-ai-app-that-undressed-women-shows-how-deepfakes-harm-the-most-vulnerable/?fbclid=IwAR0w2t6Q9i3xiCkeeZJFBOBhseWoH-mTFAh3DZVM4-8ayqWYNd6Xp6y_tGg. Acesso em: 29 jul. 2020.

HILDEBRANDT, Mireille; VRIES, Katja de (Org.). *Privacy, due process and the computational turn: the philosophy of law meets the philosophy of technology*. Abingdon, Oxon, [England]; New York: Routledge, 2013.

HOFMANN, Jeanette. Mediated democracy – Linking digital technology to political agency. *Internet Policy Review*, v. 8, n. 2, 30 jun. 2019.

KIETZMANN, Jan et al. Deepfakes: Trick or treat? *Business Horizons*, v. 63, n. 2, p. 135-146, 2020. Disponível em: <https://doi.org/10.1016/j.bushor.2019.11.006>. Acesso em: 20 ago. 2020.

MAGRANI, Eduardo. *Democracia conectada: a internet como ferramenta de engajamento político-democrático*. Curitiba: Juruá, 2014.

MARAS, Marie-Helen; ALEXANDROU, Alex. Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, v. 23, n. 3, 2019, p. 255-262. Disponível em: <https://journals.sagepub.com/doi/10.1177/1365712718807226>. Acesso em: 20 ago. 2020.

MEYER, Robinson. The Grim Conclusions of the Largest-Ever Study of Fake News. *The Atlantic*. Disponível em: <https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104/>. Acesso em: 19 jun. 2021.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. Inteligência Artificial e a Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de *machine learning*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Org.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019.

OLIVEIRA, Samuel R de. *Sorria, você está sendo filmado!:* repensando direitos na era do reconhecimento facial. São Paulo: Thomson Reuters Brasil, 2021.

PARISER, Eli. *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press, 2011.

_____. *O filtro invisível: o que a internet está escondendo de você*. Rio de Janeiro: Zahar, 2012.

RAIS, Diogo; SALES, Stela Rocha. *Fake news, deepfakes e eleições*. In: RAIS, Diogo (Coord.). *Fake news: a conexão entre desinformação e o Direito*. 2. ed. São Paulo: Thomson Reuters Brasil, 2020. Kindle Edition.

RODOTÀ, Stefano. *A vida na sociedade da vigilância: a privacidade hoje*. Rio de Janeiro: Renovar, 2008.

SCHICK, Nina. *Deepfakes: The Coming Infocalypse*. New York: Twelve, 2020.

SILBEY, Jessica M.; HARTZOG, Woodrow. *The Upside of Deep Fakes*. Rochester, NY: Social Science Research Network, 2019. Disponível em: <https://papers.ssrn.com/abstract=3452633>. Acesso em: 23 jan. 2021.

SOLOVE, Daniel. J. *The future of reputation: gossip, rumor, and privacy on the Internet*. New Haven: Yale Univ. Press, 2007.

SOUZA, Carlos Affonso. Decreto do Bolsonaro para redes sociais é absurdo e aqui vão 5 provas disso. Disponível em: <https://www.uol.com.br/tilt/colunas/carlos-affonso-de-souza/2021/06/01/decreto-de-bolsonaro-pode-criar-vale-tudo-na-internet.htm>. Acesso em: 30 jun. 2021.

SUNSTEIN, Cass. R. *Republic.com 2.0*. Princeton: Princeton University Press, 2007.

_____. *#Republic: divided democracy in the age of social media*. Third printing, and first paperback printing ed. Princeton Oxford: Princeton University Press, 2018.

SWIFT, Jonathan. No. 14. *The Examiner*, Londres, 11 set. 1710.

TEYSSOU, Denis; SPANGENBERG, Jochen. Video Verification: Motivation and Requirements. In: MEZARIS, Vasileios et al. (Ed.). *Video Verification in the Fake News Era*. Cham: Springer, 2019.

TRUSSLER, Marc; SOROKA, Stuart. Consumer Demand for Cynical and Negative News Frames. *The International Journal of Press/Politics*, v. 19, n. 3, p. 360-379, 2014.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. *Science*, v. 359, n. 6380, p. 1146-1151, 2018.

WESTERLUND, Mika. The Emergence of Deepfake Technology: A Review. In: *Technology Innovation Management Review*, [s.l.], nov. 2019. Disponível em: <https://timreview.ca/article/1282>. Acesso em: 29 jul. 2020.

WOJEWIDKA, John. The deepfake threat to face biometrics. *Biometric Technology Today*, v. 2, n. 2, fev. 2020, p. 1-12. Disponível em: <https://www.sciencedirect.com/journal/biometric-technology-today/vol/2020/issue/2>. Acesso em: 20 ago. 2020.

WU, Tim. *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Vintage, 2017.

Sobre a autora e o autor:**Caitlin Mulholland** | *E-mail:* caitlin.mulholland@gmail.com

Doutora e Mestre em Direito Civil pela Universidade do Estado do Rio de Janeiro (2006 e 2002). Professora Associada do Departamento de Direito da Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), onde atualmente coordena a graduação em Direito. Coordenadora do Mestrado Profissional em Direito Civil Contemporâneo e Prática Jurídica da PUC-Rio. Professora do Programa de Pós-graduação em Direito Constitucional e Teoria do Estado da PUC-Rio. Professora pesquisadora do Núcleo Legalite da PUC-Rio – Direitos e Novas Tecnologias.

Samuel Rodrigues de Oliveira | *E-mail:* samuelrdeoliveira@gmail.com

Doutorando em Teoria do Estado e Direito Constitucional pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Mestre em Direito e Inovação (2020) e bacharel em Direito (2017) pela Universidade Federal de Juiz de Fora (UFJF). Especialista em Relações Internacionais. Pesquisador no Núcleo Legalite PUC-Rio, no Instituto Legal Grounds e no Centro de Justiça e Sociedade da FGV Direito Rio.

Data da submissão: 30 de junho de 2021.

Data do aceite: 18 de outubro de 2021.