SOUZA, Carlos Eduardo; SANTAELLA, Lucia. Deepfakes na perspectiva da semiótica. *TECCOGS* – Revista Digital de Tecnologias Cognitivas, n. 23, jan./jun. 2021, p. 26-44.

Recebido em: 10 fev. 2021
Aprovado em: 1 mar. 2021

dx.doi.org/10.23925/1984-3585.2021i23p26-44

Deepfakes na perspectiva da semiótica

Carlos Eduardo de Souza[1]

Lucia Santaella[2]

Resumo: As notícias falsas são uma arma de desinformação conhecida, capaz de ameaçar o estado democrático. Em um momento em que as mídias tradicionais são constantemente atacadas e acusadas de fazerem parte de uma grande conspiração para manter o poder das classes dominantes, as pessoas transformaram as redes sociais em sua fonte primária de informação, pois nelas existe menos controle sobre o que circula e, pretensamente, trazem menor risco de manipulação pelas mídias tradicionais. Na intersecção de fatores como a democratização na produção de conteúdo sem qualquer supervisão, a personalização das mensagens que não confrontam o usuário e o estímulo ao compartilhamento, as fake news encontram um campo para florescer e se propagar. Em 2017 elas evoluíram, com o nome de deepfake, deixaram de ser apenas mensagens no formato de texto, para contar com a manipulação de imagens, áudios e vídeos. Sua capacidade de forjar a realidade de maneira praticamente imperceptível, até mesmo para especialistas, chamou a atenção das mídias e da academia. Para discutir os efeitos e as ameaças dessa tecnologia, e como combatê-las, esse artigo, baseado na semiótica desenvolvida por Peirce, aponta para o esforço feito pelos produtores do vídeo *In Event of Moon Disaster* em criar um conteúdo educativo para alertar as pessoas sobre as consequências das deepfakes.

Palavras-chave: Deepfake. Fake News. Semiótica. Peirce. Redes Sociais.

1 Carlos Eduardo de Souza é bacharel em Administração pela Universidade Presbiteriana Mackenzie e mestrando em Tecnologias da Inteligência e Design Digital, PUC-SP. CV Lattes: lattes.cnpq.br/4424563031670368. E-mail: cadu.souza81@gmail.com.

2 Lucia Santaella é pesquisadora 1A do CNPq, professora titular da PUC-SP. Publicou 51 livros e organizou 24, além da publicação de mais de 400 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011 e 2014), o prêmio Sergio Motta (2005) e o prêmio Luiz Beltrão (2010). ORCID: orcid.org/0000-0002-0681-6073. CV Lattes: lattes.cnpq.br/7427854657719431. E-mail: lbraga@pucsp.br.

Deepfakes in a semiotic perspective

Abstract: Fake news is a known weapon of disinformation capable of threatening democracy. At a time when traditional media are constantly attacked and accused of being part of a major conspiracy to maintain the power of the ruling classes, people have turned social networks into their primary source of information. As there is less control over what circulates there, it is supposed that they bring less risk of manipulation by traditional media. At the intersection of factors such as the democratization of content production without any supervision, the personalization of messages that do not confront the user and the encouragement of sharing, fake news finds a field to flourish and spread. In 2017, under the name of deepfake, they evolved from being just messages in text format, to relying on the manipulation of images, audio and videos. Its ability to forge reality practically imperceptibly, even for specialists, caught the attention of the media and academia. To discuss the effects and threats of this technology, and how to combat them, this article, based on the semiotics developed by Peirce, points to the effort made by the producers of the video "In Event of Moon Disaster" to create educational content to alert people about the consequences of deepfakes.

Key words: Deepfake. Fake News. Semiotics. Peirce. Social networks.

Fake news

No contexto da globalização, as corporações de mídias enfrentam desafios que atingem sua base estrutural, de formas e de conteúdo. Nesse cenário, acadêmicos e jornalistas, entre outros profissionais, são frequentemente confrontados com o fenômeno das notícias falsas (ou fake news) (BERDUYGINA; VLADIMIROVA; CHERNYAEVA, 2019). Em 2017, o termo "fake news" foi aclamado como termo do ano pelo dicionário Collins, seguindo o que o dicionário chamou de "presença onipresente" nos últimos doze meses então transcorridos. Segundo o monitoramento feito pelos lexicógrafos que trabalham para o Collins, a presença do termo cresceu 365% na comparação ano contra ano. Ainda de acordo com o trabalho do Collins, fake news, tem sido utilizado nos Estados Unidos para descrever "falsas, frequentemente sensacionalistas, informações disseminadas sob o pretexto de reportagem informativa" (BROWN, 2017; FLOOD, 2017, p. 1). De acordo com Santaella:

> Notícias falsas costumam ser definidas como notícias, estórias, boatos, fofocas ou rumores que são deliberadamente criados para ludibriar ou fornecer informações enganadoras. Elas visam influenciar as crenças das pessoas, manipulá-las politicamente ou causar confusões em prol de interesses escusos. (SANTAELLA, 2018, p. 263-265)

A definição apresentada pode abarcar uma grande variedade de formas de desinformação para fins comerciais e de publicidade, frequentemente com forte apelo visual (BERDUYGINA; VLADIMIROVA; CHERNYAEVA, 2019; SANTAELLA 2018). O contexto atual das fake news é substancialmente diferente daquele comumente encontrado no domínio dos meios tradicionais de comunicação, quando a produção de notícias era limitada e confiável, na medida em que seguia um conjunto de normas e princípios adotados pelos jornalistas. As novas formas de produzir e consumir informação e notícias não guardam padrões semelhantes.

Assim, o termo fake news passou a se referir a postagens virais baseadas em contas fictícias feitas para se parecerem com notícias. Vários tipos de notícias falsas estão identificados à medida que crescem os estudos sobre o tema (BOTHA; PIETERSE, 2020). Alguns dos tipos mais comuns são: (a) sátira ou paródia criadas para fins de entretenimento, sem

intenção de causar dano, mas potencialmente podendo enganar a audiência; (b) fabricação de notícias, podendo envolver manipulação de fotos ou vídeos, que são notícias não baseadas em fatos, porém publicadas como se fossem reportagens para transmitir legitimidade com o intuito de enganar; (c) manipulação de fotos, envolvendo a manipulação de imagens ou vídeos reais para estabelecer uma narrativa falsa; (d) conexão falsa, quando manchetes, imagens e/ou legendas incluem notícias ou artigos que contêm conteúdo genuíno e preciso, mas fazem uso de títulos enganosos ou sensacionalistas; (e) publicidade e relações públicas, em que a publicidade e os comunicados de imprensa são publicados como notícias, sendo muitas vezes um conteúdo patrocinado; (f) captura de cliques, notícias fabricadas propositalmente para ganhar mais visitantes em um site e aumentar a receita de publicidade; (g) conteúdo enganoso, transmitindo informações enganosas para enquadrar indivíduos ou questões (BORGES, GAMBARATO, 2019; BOTHA, PIETERSE, 2020).

Para os fins deste artigo, seguiremos com o conceito de fake news apresentado por Allcott e Gentzkow (2017, p. 2013), "artigos de notícias que são intencionalmente e comprovadamente falsos, e podem enganar os leitores". O uso de manchetes sensacionalistas para seduzir a audiência é uma prática de longa data. Entretanto, dentro das redes sociais, tanto o alcance, quanto o efeito da disseminação de conteúdo ocorrem em uma escala muito mais rápida, de modo que essa informação distorcida, imprecisa ou falsa adquire um enorme potencial para causar impactos reais, em poucos minutos (FIGUEIRA; OLIVEIRA, 2017).

De fato, a informação falsa espalha-se rapidamente pelas redes sociais, podendo impactar milhões de usuários (ibid.), como sugerem dois estudos conduzidos por Allcott e Gentzkow. No primeiro, constatou-se que 38 milhões de compartilhamentos de notícias falsas que aconteceram nas redes sociais, resultaram em 760 milhões de ocorrências de um usuário clicando e lendo uma notícia falsa, ou cerca de três histórias lidas por adulto americano. No segundo, uma lista de sites de notícias falsas, na qual pouco mais da metade dos artigos parecem ser falsos, recebeu 159 milhões de visitas durante o mês da eleição, ou 0,64 por adulto nos EUA (ALLCOTT; GENTZKOW, 2017).

De acordo com Westerlund (2019), atualmente, 20% dos usuários da Internet obtêm suas notícias via YouTube, sendo essa porcentagem menor apenas que a de usuários que obtém informações pelo Facebook. A crescente popularidade do vídeo mostra a importância da criação de ferramentas que confirmem a autenticidade do conteúdo dessa mídia e notícias, pois à medida que as novas tecnologias permitem a alteração convincente do conteúdo, torna-se mais fácil obter e divulgar informações incorretas através das mídias sociais.

À medida que o crescimento das redes sociais vem destruindo as barreiras de entrada que existiam para prevenir a disseminação das fakes news, esse fenômeno permite que qualquer pessoa possa criar e disseminar conteúdo (BERDUYGINA; VLADIMIROVA; CHERNYAEVA, 2019). Suportada pela lógica das redes sociais e dos buscadores de informação, de facilidade de publicação e compartilhamento de conteúdo sem qualquer avaliação de terceiros, sem checagem dos fatos ou critério editorial, qualquer conteúdo produzido pode atingir milhões de usuários (ALLCOTT; GENTZKOW, 2017). O ápice dessa lógica parece ter sido atingido pela popularização das mídias móveis, que tornaram qualquer lugar um ponto de produção e compartilhamento instantâneo de informação. Imagem, som e vídeo podem ser criados e disseminado por milhões de pessoas para milhões de pessoas em diversas plataformas, por usuários que muitas vezes desconhecem o funcionamento dos algoritmos que funcionam nessas redes (SANTAELLA, 2018).

A partir do momento em que cresce a descrença nas mídias tradicionais, fica aberto o espaço para que as mídias alternativas as desqualifiquem como não confiáveis. Ademais, à medida que as notícias falsas tornam-se cada vez mais sofisticadas, criam-se as condições para o império da "pós-verdade", que é caracterizado pela desinformação digital e guerra de informação liderada por atores malévolos executando campanhas de informações falsas para manipular a opinião pública e aumentar a polarização política, usando para isso canais de comunicação alternativos, independentes e descentralizados sem qualquer compromisso com a informação factual (WESTERLUND, 2019).

A impulsão dessas informações conta com diversas heurísticas cognitivas que compõem três fenômenos em particular – a dinâmica da "cascata de informações", a atração humana por informações negativas e novas, e as bolhas de isolamento especialmente úteis para explicar por que notícias se tornam virais (CHESNEY; CITRON, 2019). A cascata de informações está relacionada com a dinâmica que surge a partir do momento em que as pessoas não prestam atenção nas informações que estão compartilhando, pois presumem que outros determinaram de forma confiável a credibilidade da informação antes de transmiti-la (ibid.). Sendo levadas a questionar toda a informação que recebem das mídias tradicionais, as pessoas se protegem usando como fontes confiáveis suas redes sociais e buscam opiniões que apoiem suas ideias existentes. Na verdade, muitas pessoas estão abertas a qualquer coisa que confirme suas visões existentes, mesmo que suspeitem que seja falso (WESTERLUND, 2019).

A lógica dos algoritmos de recomendação tem um papel muito relevante no contexto das notícias falsas, possivelmente criando bolhas de isolamento, uma vez que o algoritmo assume como verdade a predileção do usuário por determinado conteúdo, assim como cria câmaras de eco, já que há um maior peso para conteúdo publicado e compartilhado por pessoas que declaram ter as mesmas orientações do usuário, gerando, assim, um isolamento contra opiniões contrárias (BORGES; GAMBARATO, 2019; CHESNEY; CITRON, 2019).

Pesquisas demonstram que as pessoas frequentemente ignoram informações que contradizem suas crenças e interpretam evidências ambíguas como consistentes desde que alinhadas com as suas crenças (CHESNEY; CITRON, 2019). A consequência desse fenômeno é a fragmentação da "esfera pública" em subesferas fortemente separadas, cada uma delas fluindo de acordo com sua própria interpretação do "mundo autocongruente, geralmente intolerante e agressiva para as demais subesferas, com as quais a possibilidade de o diálogo e a compreensão diminuiriam rapidamente" (CITTON, 2021, p. 49).

Outra característica das redes sociais é sua operação com base na quantidade de cliques e volume de tráfico que uma determinada publicação recebe, uma vez que não há juízo de valor, basta a menor interação com o conteúdo falso para que o algoritmo contabilize esse fato como interesse naquele conteúdo (SANTAELLA, 2018). A partir do momento em que essa interação é capturada, o processo de personalização do conteúdo se encarrega de perpetuar os ecos que passam através dos filtros da bolha criada (BORGES; GAMBARATO, 2019). Todo o processo está baseado em grandes bancos de dados, nos quais são armazenadas informações sobre quais sites o usuário acessou, com que perfis interagiu e de que forma, o que foi compartilhado; praticamente todas as interações feitas através do computador são registradas e utilizadas para selecionar as informações que o usuário receberá. Esse grau de personalização pode ter consequências na forma como o usuário avalia o mundo e toma decisões (GUARDA; OHLSON; ROMANINI, 2018).

Conforme praticamente todas as nossas interações na internet são coletadas, agrupadas e analisadas por algoritmos de Inteligência Artificial (IA), as redes sociais e buscadores de informação passam a controlar e moldar nosso consumo de informação. Assim, a IA torna-se extremamente capaz de prever o que nos seduz, vicia ou enfurece, para manter-nos clicando, lendo, assistindo e compartilhando. Ademais, essa mesma IA é capaz de forjar mídias sintéticas em tempo real calibrado para explorar esses vieses (FLETCHER, 2018). Nessa intersecção, reside o combustível para o pesadelo do fim das democracias.

Ao fim e ao cabo, o poder das notícias falsas provém do apelo ao emocional, que captura rapidamente a atenção do leitor, levando-o, além do clique, ao compartilhamento, apenas com base na manchete, sem qualquer filtro de criticidade ao conteúdo da postagem (SANTAELLA, 2018). No modelo de análise das redes sociais como fonte de informação desenvolvido por Allcott e Gentzkow (2017, p. 221), também encontramos elementos que demonstram a permissividade dessa plataformas, (a) baixos custos de entrada e produção de conteúdo, o que torna as estratégias de produção de notícias falsas bastante rentáveis; (b) o formato em que as postagens são exibidas torna difícil avaliar a veracidade de um artigo; (c) nas redes sociais as pessoas estão mais propensas a ler o que pessoas com a mesma orientação ideológica publicam e compartilham.

Santaella (ibid.), destaca três traços das notícias falsas que instigam sua propagação pela internet, "desinformação, desconfiança e manipulação". Considerando que atravessamos um momento de persistência da "pós verdade", termo que a literatura psicológica trata como viés de confirmação, ou seja, a tendência que os indivíduos têm em praticar uma escuta seletiva (ROMANINI; OHLSON, 2018), a relação dos usuários nas mídias sociais é mais influenciada pelas emoções do que pela criticidade quanto ao conteúdo adulterado, que ganha cada vez mais espaço para crescimento, uma vez que as emoções e crenças pessoais tornam-se mais importantes na formação da opinião pública do que os fatos verificados (GUARDA; OHLSON; ROMANINI, 2018).

Nesse contexto, as revelações das pesquisas em psicologia social e outros campos aumentam as preocupações quanto ao consumo de informação, pois ficam minadas as antigas teorias de humanos como agentes que avaliavam friamente as evidências frente aos seus valores e objetivos, para então modificar essas crenças em respostas às evidências. Tais pesquisas apontam o contrário, para o fato de que primeiro adquirimos crenças e apegos, e que posteriormente filtramos as evidências, muitas vezes inconscientemente, de forma a manter a coerência das nossas identidades (FLETCHER, 2018).

> Nós alinhamos nossas lentes de crença / interpretação com as de nosso grupo social (tanto online quanto off-line), explicando evidências e interpretações que nos alienariam do "nós" do nosso grupo. Da mesma forma, detectamos e evitamos evidências e interpretações de que se identifique com "eles". (FLETCHER, 2018, p. 466)

Os relatos das notícias falsas excluem a realidade externa ou a distorcem propositalmente. Separar notícias falsas de notícias verdadeiras baseia-se em conhecer a relação entre a reportagem, o fato externo ao qual ela se refere e qual a intenção por trás da reportagem. Evidentemen-

te a linha entre o uso de processos equivocados e ações deliberadamente enganosas não é clara, e a verdade não é intrínseca à notícia; a esta cabe o papel de narrar um acontecimento (BORGES; GAMBARATO, 2019). Sob esse aspecto, a semiótica tem uma contribuição a dar, conforme será discutido adiante. Por ora, é importante apontar para a intensificação dos problemas quando as fake news são convertidas em deepfakes.

Das fake news às deepfakes

O aspecto histórico das fake news não é algo novo e há razões para crer que o papel desse tipo de conteúdo continuará a crescer. Com menores barreiras de entrada, incentivos a monetização de conteúdo on-line, queda na confiança nos meios tradicionais de comunicação em massa, crescimento das mídias sociais e o aumento da polarização política (ALLCOTT; GENTZKOW, 2017), novas formas de fake news. Surgiram com potencial ainda maior de enganar a audiência, como são as deepfakes.

A primeira aparição da tecnologia deepfake foi na plataforma de mídia social *Reddit* publicada por um usuário anônimo em novembro de 2017 (BOTHA; PIETERSE, 2020). As deepfakes podem ser definidas como mídias sintéticas geradas com o uso de IA. É uma junção das expressões *deep learning* (aprendizado profundo) e *fake* (falso) (ASHISH, 2020; BATTAGLIA, 2020). Deepfakes também podem ser entendidas como o produto de aplicativos de IA que fundem, combinam, substituem e sobrepõem imagens e vídeo clipes para criar vídeos falsos que parecem autênticos (WESTERLUND, 2019). Para nossos propósitos, seguiremos com a definição apresentada por Chesney e Citron (2019, p. 1757-1758),

> Isso deu origem ao rótulo de "deepfakes" para essas personificações digitalizadas. Usamos esse rótulo aqui de forma mais ampla, como uma abreviatura para toda a gama de falsificações digitais hiper-realistas de imagens, vídeo e áudio. Esta gama completa implicará, mais cedo ou mais tarde, uma perturbadora gama de usos maliciosos. Não somos de forma alguma os primeiros a observar que *deep fakes* irão migrar muito além do contexto da pornografia, com grande potencial de danos.

Assim, como uma combinação de "*deep learning*" e *fake*, as deepfakes são mídias sintéticas geradas por IA, altamente realistas, de difícil detecção (CHESNEY; CITRON, 2019). A partir dessa tecnologia é possível produzir vídeos hiper-realistas manipulados digitalmente para representar pessoas dizendo e fazendo coisas que nunca realmente aconteceram. Contando com redes neurais que analisam grandes conjuntos de amostras de dados para aprender a imitar as expressões faciais, maneirismos, voz e inflexões

de uma pessoa, é possível colocar qualquer um em qualquer situação (WESTERLUND, 2019).

No passado, a manipulação de vídeos exigia grandes recursos e estava à disposição de poucas empresas. Atualmente, entretanto, computadores domésticos permitem "uma assustadoramente precisa troca de rosto em um único computador de jogo, possivelmente em menos de vinte e quatro horas. Sem equipe, sem recursos, sem dinheiro" (FLETCHER, 2018, p. 463). Essas tecnologias têm o potencial de criar falsificações mais reais e profundas, mais difíceis de detectar, com um potencial ainda maior de sabotagem à democracia, pois permitem a produção de vídeos de notícias aparentemente legítimas que põem em cheque a reputação de jornalistas e da mídia, colocam pessoas em lugares e situações nas quais elas nunca estiveram com o interesse em distorcer a opinião pública (GUARDA; OHLSON; ROMANINI, 2018; WESTERLUND, 2019). Sua eficácia e suas ameaças não estão apenas em sua capacidade de forjar realidades, "mas sim em sua capacidade de ressoar dentro do atual estado afetivo das multidões" (CITTON, 2021, p. 50).

A disponibilidade dos algoritmos utilizados para a criação de mídias sintéticas a partir de IA permitiu a rápida automação do processo de deepfake. Para criar um vídeo com conteúdo falso é necessário apenas a seleção de imagens da face da pessoa que será substituída e da face da pessoa que será sobreposta (BOTHA; PIETERSE, 2020). Com tal facilidade disponível torna-se cada mais difícil saber em que confiar, o que resulta em prejuízos para a tomada de decisão, entre outras coisas (WESTERLUND, 2019).

O ponto de inflexão das deepfakes encontra-se no escopo, escala e na sofisticação da tecnologia envolvida, já que quase qualquer pessoa com um computador pode fabricar vídeos falsos que são praticamente indistinguíveis da mídia autêntica (FLETCHER, 2018). É provável que, no futuro, as deepfakes evoluam para pornografia de vingança, *cyberbullying*, desqualificação de provas em tribunais, sabotagem política, propaganda terrorista, chantagem, manipulação de mercado e notícias falsas (WESTERLUND, 2019).

Com os avanços tecnológico, é razoável apostar em um futuro no qual a guerra da desinformação estará mais bem estruturada e o uso da IA para criar e entregar vídeos falsos adaptados aos preconceitos específicos dos usuários de mídia social estará disseminado. As chamadas deepfakes serão a desinformação transformada em armas destinadas a interferir nas eleições e semear agitação civil (ibid.).

As redes sociais e buscadores de internet, ao proporcionarem o contato frequente com a desinformação, fazem com que as pessoas não se sintam confiantes em toda a informação disponível, resultando em um fenômeno denominado de "apocalipse da informação" ou "apatia da realidade". Além disso, as pessoas podem até descartar as filmagens genuínas como falsas, simplesmente porque se enraizaram na noção de que tudo em que não querem acreditar deve ser falso. Em outras palavras, a maior ameaça não é que as pessoas sejam enganadas, mas que passem a considerar tudo como engano (ibid.).

Em um mundo onde praticamente tudo pode ser manipulado artificialmente, a "apatia da realidade" pode representar uma grave ameaça à capacidade da sociedade compartilhar e cooperar com base em "percepções precisas até mesmo das mais básicas ou urgentes realidades, tornando as sociedades ainda mais vulneráveis ao governo autocrático" (FLETCHER, 2018, p. 465).

Nessa batalha cibernética, as técnicas computacionais para detecção das deepfakes estão focadas em identificar um artefato específico, mas não generalizam bem para detectar novas versões. O desempenho dos detectores de última geração está diminuindo rapidamente à medida que a qualidade das deepfakes melhora (MIRSKY; LEE, 2021; KORSHUNOV; MARCEL, 2018). A divulgação de novos avanços no combate às deepfakes leva a uma evolução da sua produção, criando uma corrida armamentista de IA, com o aperfeiçoamento das técnicas destinadas a gerar deepfakes cada vez mais sofisticadas e difíceis de detectar. Assim, trocamos um sabor de distopia de uma ficção científica por outra (FLETCHER, 2018), à medida que cada ciclo de novas tecnologias de combate à deepfake carrega a semente da evolução da próxima geração de deepfakes.

Tais condições só aumentam a importância da mídia confiável para a preservação da democracia. Essas preocupações são mais oportunas e relevantes do que nunca: vivemos em um contexto dominado pela "pós verdade", em uma era de crescente populismo autoritário, acompanhada de repressões em veículos de jornalismo legítimo e demandas de base por justiça ambiental e racial. Urge que os cidadãos obtenham informações confiáveis e um melhor entendimento de como se envolver com este ambiente de mídia fragmentado (MOONDISASTER, 2021).

Infelizmente, a alfabetização midiática não pode fornecer um antídoto mágico para o tremendo aumento da desinformação. Como estratégia pedagógica central, no entanto, pode nos ajudar a cultivar a crítica interior, transformando-nos de consumidores passivos da mídia em um público criterioso. Ensinar sobre deepfakes significa alertar para a ameaça perniciosa que representam, para as várias abordagens que visam combatê-las e para os usos alternativos de mídia sintética (ibid.).

Assim, abordagens como a educação e o treinamento são cruciais nesse combate. É preciso aumentar a conscientização pública tanto entre nativos digitais quanto pessoas mais velhas, para compreender que um vídeo, ao contrário do que parece, pode não fornecer uma informação precisa do que aconteceu, e encontrar quais pistas perceptivas podem ajudar a identificar e combater o efeito danoso das deepfakes (WESTERLUND, 2019 MIRSKY; LEE, 2021). Um caminho auxiliar que se apresenta para isso é também o da semiótica, conforme será discutido abaixo a partir de um caso exemplar.

In event of moon disaster

O caso exemplar, a ser comentado semioticamente, é a obra "In event of moon disaster". Em uma intersecção do campo da arte e política, os pesquisadores do MIT, Francesca Panetta e Halsey Burgund, propuseram uma ideia provocativa usando a tecnologia das deepfakes, através da qual procuram mostrar o potencial da tecnologia para educar as pessoas sobre seu uso. Eles escolheram recriar uma versão alternativa da viagem à Lua, realizada pela nave Apollo em 1969 (HAO, 2020; MOONDISASTER, 2021). Antes da missão, os redatores do presidente Richard Nixon elaboraram dois discursos, um para o cenário de sucesso e outro, designado "In event of moon disaster", para caso as coisas não saíssem de acordo com o planejado. O verdadeiro Nixon, nunca precisou proferir o segundo, porém, a partir da sobreposição de imagens e áudio os pesquisadores foram capazes de representar o que seria o presidente Nixon proferindo o segundo discurso (HAO, 2020). Questionados se eles não estariam disseminando informações falsas os autores da produção foram enfáticos em negar tal pergunta:

> Como artistas multimídia e jornalistas que trabalharam por uma década em um cenário de mídia em constante mudança, acreditamos que as informações apresentadas como falsas em um contexto artístico e educacional não são desinformações. Na verdade, pode ser fortalecedor: experimentar um uso poderoso de novas tecnologias de forma transparente que tem o potencial de ficar com os espectadores e torná-los mais cautelosos sobre o que verão no futuro. Usando as técnicas mais avançadas disponíveis e insistindo na criação de um vídeo usando visuais sintéticos e áudio sintético (um "deepfake completo"), pretendemos mostrar para onde essa tecnologia está caminhando - e quais podem ser algumas das principais consequências (MOONDISASTER, 2021, p. 1).

In event of moon disaster ilustra as possibilidades de tecnologias deepfake ao recriar uma versão alternativa da missão da Apollo 11. O ponto de partida é: o que teria acontecido caso algo de errado ocorresse e os astronautas não pudessem voltar para casa? Um discurso de contingência

para essa possibilidade foi preparado, mas nunca feito pelo presidente Nixon – até agora. Nesse projeto de arte foi criada uma história alternativa, pedindo a todos nós que consideremos como as novas tecnologias podem dobrar, redirecionar e ofuscar a verdade ao nosso redor (MOONDISASTER, 2021).

Para construir essa versão alternativa da história, várias técnicas de desinformação foram usadas – desde a edição enganosa simples até tecnologias deepfakes mais complexas. Para recriar o discurso de contingência, a peça usou técnicas de *deep learning* para criar uma voz sintética de Nixon e técnicas de substituição de diálogo para replicar o movimento da boca e dos lábios de Nixon. Ao criar essa história alternativa, o projeto explora a influência e a difusão da desinformação e das tecnologias das deepfakes em nossa sociedade contemporânea (MOONDISASTER, 2021). Esse trabalho funciona como um exemplar a ser utilizado para finalidades educativas sob o olhar da semiótica.

A contribuição da semiótica

A semiótica oferece uma série de elementos para a análise da interpretação "dos signos produzidos pelas novas tecnologias, assim como do seu papel potencial em ambientes com grandes concentrações de notícias falsas, como as redes sociais" (FERRAREZI; BORGES, 2020, p. 60). Assim, a semiótica, na vertente desenvolvida por C. S. Peirce, pode servir de bússola ao longo do debate sobre o tema,

> Pois, como ciência da significação, da denotação e da interpretação dos processos de linguagem e de comunicação, essa ciência pode nos oferecer conceitos fundamentais, capazes de nos guiar na tarefa de perscrutar os modos de produção, interpretação e disseminação das Fake News. [...] Uma ciência de base filosófica e, como ciência, cria conceitos com a finalidade de nos ajudar a pensar. Portanto, para apreender esses conceitos é preciso exercitar a paciência teórica. Só isso pode trazer compensações consequentes para os modos como interpretamos os problemas relativos às Fake News. (SANTAELLA, 2020, p. 13-14).

A semiótica tem no signo seu elemento central. Sob certo aspecto ou capacidade, o signo representa algo para um intérprete. Todas as formas de comunicação podem ser consideradas signos (SANTAELLA, 2020; BORGES; GAMBARATO, 2019). O signo é composto por três elementos, o signo, o objeto e o interpretante. Sua função representativa é aquela de mediar entre o objeto representado e o efeito que produz na mente do intérprete, efeito chamado de interpretante (BORGES; GAM-

BARATO, 2019; SANTAELLA, 2020; FERRAREZI; BORGES, 2020). À medida que os signos representam alguma coisa, ou seja, seu objeto em alguns de seus aspectos, com referência a algum um tipo de ideia, eles são, portanto, mais ou menos precisos e também podem ser usados para enganar, quando existe uma colisão em vez de uma correspondência entre o signo e aquilo que ele professa representar (BORGES; GAMBARATO, 2019).

Entretanto, o intérprete não está limitado à representação que um determinado signo faz do seu objeto. Sendo os signos por natureza incompletos, isso permite ao intérprete reportar-se ao contexto do objeto do signo por meio da experiência colateral que teve, tem ou poderá vir a ter com ele" (SANTAELLA, 2020, p. 16). Segundo esta autora (ibid., p. 18):

> Se o signo é parte de um contexto existencial, factual, maior do que ele, sua verdade ou falsidade pode ser averiguada por experiência colateral com o objeto do signo, quer dizer, o campo de referências do signo. Isso é justamente aquilo a que Hanna Arendt (1972) deu o nome de verdade factual, que é, efetivamente, a única classe de signo que, pelo fato de funcionar como um indicador, um índice de seu objeto de referência, a saber, o acontecimento, o fato ocorrido, pode ser interpretado como verdadeiro ou falso, por meio do rastreamento desse objeto de referência. Essa distinção sígnica precisa ser feita para se evitar que tudo, indiscriminadamente, entre no saco de gatos das Fake News.

Assim, a semiótica pode ser empregada como modelo de análise das deepfakes, e, em especial, do experimento conduzido por Francesca Panetta e Halsey Burgund, acima descrito, na medida em que esse caso pode contribuir para criar processos educativos que auxiliem as pessoas a repensarem suas visões do mundo, do outro e de si mesmos. Uma vez que fake news e deepfakes são inegavelmente signos, do tipo factuais, ou seja, representam fatos, elas implicam as noções de realidade (a realidade dos fatos) e de verdade (dentro de suas capacidades e limites o signo pode apresentar fidelidade aos fatos ou, então, mentir em relação a eles). Como se pode ver, os conceitos de realidade e verdade estão implícitos e decorrem da noção de signos (ROMANINI; OHLSON, 2018). Segundo NÖTH (2016, p. 137), o real se define como

> Independente dos meus e dos seus caprichos (CP 5.311). Em 1903, ele postula que o real é como é, independentemente de como imaginamos que ele seja (1903, CP 7.659). Também em 1877, Peirce define o real como independente de qualquer conhecedor, [...] o real Peirciano não é apenas um *ens*, um modo de ser; ele age sobre nossos sentidos. Existem coisas Reais cujas características são inteiramente independentes de nossas opiniões sobre elas.

Embora a realidade nela mesma seja algo que independe do que possamos pensar sobre ela, a realidade nos é acessível pela mediação do signo, podendo, assim, afetar nossos pensamentos, produzindo como efeito uma ideia, ou melhor, um interpretante coincidente ou não com a realidade. Assim, a contribuição da semiótica para refletir sobre o combate às fake news e deepfakes, reside na conexão necessária entre o pensamento e a realidade (BORGES; GAMBARATO, 2019).

Ora, signos são por natureza sociais. Portanto, na perspectiva de Peirce, os acontecimentos sociais estão conectados à experiência humana, e, logo, devem ser analisados em um processo lógico de investigação comunitário, interessado na verdade dos fatos. É no confronto dos nossos julgamentos isolados com o julgamento da comunidade que surge a verdade factual. Contudo, jamais estamos de pose da última verdade, no máximo nos esforçamos para atingir um estágio de crença, a partir do uso de recursos materiais e tempo, no qual nos sentimos confortáveis e não encontramos benefícios práticos em prosseguir para um grau maior de precisão, com mais investigação (ROMANINI; OHLSON, 2018).

De acordo com Peirce, há quatro métodos através dos quais atinge-se a fixação de uma ideia e se estabelece uma crença. Três oferecem o conforto da crença fácil, ao passo que limitam a busca da verdade: o método da tenacidade, no qual por afinidade o indivíduo se apega a uma crença e nega qualquer evidência que a confronte, permanecendo em seu estado de conforto; o método dogmático, quando uma instituição passa a ter o poder de determinar o que é verdade e justificar a crença; e o método a priori, quando o indivíduo assume como verdadeiro um sistema de proposições universais e passa a aceitar apenas os fatos que confirmam essas proposições. Esses métodos, combinados ou não, são a base das estratégias de produção e distribuição de fake news e deepfakes (ROMANINI; OHLSON, 2018; GUARDA; OHLSON; ROMANINI, 2018; BORGES; GAMBARATO, 2019). O último método é chamado científico e, segundo Peirce, nele nossas crenças são formadas pela aceitação de algo externo, não sendo influenciadas por nossas próprias fantasias, mas sim por eventos externos mediados por signos confiáveis. Uma crença comum baseada na força de eventos externos e compartilhada por muitos pode ser chamada de chamada de conclusão (BORGES; GAMBARATO, 2019).

É sob esse aspecto que a produção *In event of moon disaster* pode ser utilizada como um caso exemplar para o desenvolvimento de contra crenças no combate às deepfakes, em razão do didatismo com que nos apresenta a facção de uma deepfake e dos efeitos de credulidade que elas estão aptas a produzir. Cabe muito bem aqui a sugestão de McLuhan (1964, p. 66 *apud* BIGGIO; BUSTAMANTE, 2021, p. 161):

> A capacidade do artista de se desviar do golpe violento da nova tecnologia em qualquer época, e impedir tal violência com plena consciência, é antiquíssimo [...]. O artista pode corrigir as relações dos sentidos antes que o golpe da nova tecnologia tenha entorpecido os procedimentos conscientes. Ele pode corrigi-los antes que o entorpecimento e a tentativa subliminar e a reação comecem.

Ao alertar sobre os perigos das novas tecnologias usadas na produção de mídias sintéticas, os autores contribuem para o "enriquecimento dos interpretantes que podem ser gerados a partir do cotejo cuidadoso das relações entre o signo e aquilo a que ele se refere" (SANTAELLA, 2020, p. 22).

Considerações finais

O uso de notícias falsas como fonte de desinformação não é um fenômeno inédito na história. O que está em rápida transformação são os usos sem precedentes de dados, algoritmos e uma infraestrutura de comunicação global com capacidade transformar qualquer pessoa em produtor de conteúdo com potencial de atingir milhões de pessoas em pouco tempo.

A democratização dos meios para a produção de conteúdo tirou dos antigos conglomerados de mídia de massa a primazia de fonte de informação e a transferiu para as redes sociais. Longe dos métodos tradicionais de certificação de produção de conteúdo, impulsionados pelos baixos custos de entrada e possibilidade rápida de rentabilizar o conteúdo por meio da receita de propaganda, esses espaços tornaram-se prolíficos para a proliferação de conteúdo falso.

Impulsionados pela lógica do compartilhamento, dos bancos de dados com registros de toda interação on-line dos usuários e algoritmos capazes de entregar informação extremamente personalizada aos usuários, que cada vez mais tem nas redes e buscadores sua fonte principal de informação, as notícias falsas se alastraram e mostraram-se capazes de decidir os destinos das sociedades.

Um passo significativo foi dado em 2017 com o surgimento das deepfakes, que ao se apoiarem na imensa quantidade de imagens e vídeos disponíveis na rede e usando o compartilhamento dos programas de código aberto, colocaram a poucos cliques de distância de qualquer pessoa com um computador a possibilidade da criação de vídeos, que vão desde a sátira e paródia com a finalidade de rir, passando pelo pornô de vingança, *cyberbullying* e manipulação da opinião pública.

No entanto a maior ameaça das deepfakes está, além da sua capacidade de colocar pessoas em lugares e situações nas quais elas nunca estiveram ou em dizer coisas que elas nunca disseram ou ainda sua difícil detecção, a maior ameaça reside na possibilidade de levar as pessoas a simplesmente duvidar de tudo o que veem, a constante exposição a conteúdos falsos pode tornar as sociedades incapazes de compartilhar e cooperar com base em observações básicas da realidade. O combate a essa ameaça é urgente e, até o momento, o que os estudiosos têm encontrado é a possibilidade de entramos em uma guerra sem fim, pois, a cada nova geração de ferramentas para combater as deepfakes, as bases para que elas sigam evoluindo já estariam lançadas.

A educação dos cidadãos sobre essa ameaça e o desenvolvimento de um senso crítico sobre as notícias que circulam na mídia podem funcionar como antídotos. A produção "*In the event of moon disaster*", ao recriar um hipotético discurso proferido pelo ex-presidente americano Richard Nixon, procura alertar os incautos para as possibilidades dessa tecnologia. Sendo uma questão que remete necessariamente ao que deveríamos entender por realidade e verdade, encontramos na semiótica de Peirce um arcabouço capaz de acompanhar a leitura dos impactos desse fenômeno. Ainda que este artigo não tenha investigado o contexto cultural, econômico e político, a saber, o contexto dos objetos dos signos deepfakes, como elementos que afetam a fixação das crenças, o pragmatismo de Peirce nos aponta um caminho promissor para conhecer as estratégias que os produtores desse tipo de conteúdo empregam e, consequentemente, formas de combatê-lo, especialmente por meio de alertas importantes para os impactos causados pelo consumo e compartilhamento de notícias e vídeos sem a análise crítica do que está por trás deles. E o que está por trás é sempre mais alarmante do que podemos imaginar. Há, portanto, que enfrentá-lo, cara a cara.

Referências

ALLCOTT, Hunt; GENTZKOW, Matthew. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, v. 31, n. 2, p. 211-236, 2017.

ASHISH, Jaiman. AI generated synthetic media, aka deepfakes. *Towards Data Science*, 9 ago., 2020. Disponível em: towardsdatascience.com/ai-generated-synthetic-media-aka-*deepfakes*-7c021dea40e1. Acesso em: 13 jul. 2021.

_____. Positive use cases of deepfakes. *Towards Data Science*, 14 ago. 2020. Disponível em: towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387. Acesso em: 13 jul. 2021.

BATTAGLIA, Rafael. Afinal, o que são deepfakes? *Super Interessante*, 29 out. 2020. Disponível em: super.abril.com.br/tecnologia/afinal-o-que-sao-*deepfakes*. Acesso em: 15 jul. 2021.

BERDUYGINA, Oksana N.; VLADIMIROVA, Tatyana N.; CHERNYAEVA, Elena V. Trends in the spread of fake news in mass media. *Media Watch*, v. 10, n. 1, p. 122-132, 2019.

BIGGIO, Federico; BUSTAMANTE, Victoria Vanessa dos Santos. Elusive masks: A semiotic approach of contemporary acts of masking. *Lexia: Revista de Semiótica*, v. 37-38, p. 141-164, 2021.

BORGES, Priscila Monteiro; GAMBARATO, Renira Rampazzo. The role of beliefs and behavior on Facebook: A semiotic approach to algorithms, fake news, and transmedia journalism. *International Journal of Communication*, v. 13, p. 603-618, 2019.

BOTHA, Johnny; PIETERSE, Heloise. Fake news and deepfakes: A dangerous threat for 21st century information security. *Anais* ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited, 2020, p. 57.

BROWN, Mike. Why 'fake news' won Collins dictionary's word of the year. *Inverse*, 11 fev. 2017. Disponível em: inverse.com/article/38041-donald-trump-*fake-news*-word-of-the-year. Acesso em: 15 jul. 2021.

CHESNEY, Bobby; CITRON, Danielle. Deep *fakes*: A looming challenge for privacy, democracy, and national security. *California Law Review*, v. 107, n. 6, p. 1753-1820, 2019.

CITTON, Yves. Could deep *fake*s uncover the deeper truth of an ontology of the networked images? *The Nordic Journal of Aesthetics*, v. 30, n. 61-62, p. 46-64, 2021.

DAVE, Johnson. What is a deep*fake*? Everything you need to know about the AI-powered *fake* media. *Insider*, 22 jan. 2021. Disponível em: businessinsider.com/what-is-deep*fake*. Acesso em: 13 jul. 2021.

FERRAREZI, Fernanda; BORGES, Priscila. O que é e o que parece ser: Imagens criadas por inteligência artificial como elementos atuantes na pós-verdade. *Anais*... 43º Congresso Brasileiro de Ciências da Comunicação Virtual. INTERCOM – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação e Universidade Federal da Bahia, 1-10 dez. 2020, p. 388-416.

FIGUEIRA, Álvaro; OLIVEIRA, Luciana. The current state of fake news: Challenges and opportunities. *Procedia Computer Science*, v. 121, p. 817-825, 2017.

FLETCHER, John. Deepfakes, artificial intelligence, and some kind of dystopia: The new faces of online post-fact performance. *Theatre Journal*, v. 70, n. 4, p. 455-471, 2018.

FLOOD, Alison. Fake news is 'very real' word of the year for 2017. *The Guadian*, 02 nov. 2017. Disponível em: theguardian.com/books/2017/nov/02/*fake-news*-is-very-real-word-of-the-year-for-2017. Acesso em: 15 jul. 2021.

GUARDA, Rebeka F.; OHLSON, Marcia P.; ROMANINI, Anderson V. Disinformation, dystopia and post-reality in social media: A semiotic-cognitive perspective. *Education for Information*, v. 34, n. 3, p. 185-197, 2018.

HAO, Karen. Inside the strange new world of being a deepfake actor. *MIT Technology Review*, 09 out. 2020. Disponível em: technologyreview.com/2020/10/09/1009850/ai-deepfake-acting/. Acesso em: 15 jul. 2021.

KORSHUNOV, Pavel; MARCEL, Sébastien. Deepfakes: A new threat to face recognition? Assessment and detection. *Idiap (=Istituto Dalle Molle di Intelligenza Artificiale Percettiva) Research Report* 18-2018, Martigny, 2018. Disponível em: publications.idiap.ch/downloads/reports/2018/Korshunov_Idiap-RR-18-2018.pdf. Acesso em: 15 jul. 2021.

MIRSKY, Yisroel; LEE, Wenke. The creation and detection of deepfakes. *ACM Computing Surveys*, v. 54, n. 1, p.1-41, 2021.

MOONDISASTER. Resources. Disponível em: moondisaster.org/about. Acesso em: 15 jul. 2021.

NÖTH, Winfried. Reconstruções semióticas da realidade: Reflexões sobre a realidade puramente objetiva de John Deely. *TECCOGS: Revista de Tecnologias Cognitivas*, n. 13, p. 132-140, 2016.

PEIRCE, Charles S. Fraser's "The Works of George Berkeley". *North American Review*, v. 113, p. 449-472, out. 1871.

ROMANINI, Anderson Vinicius; OHLSON, Márcia Pinheiro. De elos bem fechados: O pragmatismo e a semiótica peirceana como fundamentos para a tecnologia blockchain utilizada no combate às fake news. *Revista Communicare*, São Paulo, v. 18, n. 2, p. 60-73, 2018.

SANTAELLA, Lucia. *A pós-verdade é verdadeira ou falsa?* (Coleção Interrogações). Barueri: Estação das Letras e Cores, 2018.

_____. A semiótica das fake news. *Verbum – Cadernos de pós-graduação*, v. 9, n. 2, p. 9-25, 2020.

WAAL, Cornelis de. *Peirce*: A Guide for the Perplexed. London: Bloomsbury, 2013.

WESTERLUND, Mika. The emergence of deepfake technology: A review. *Technology Innovation Management Review*, v. 9, n. 11, p. 39-52, 2019.