

Deepfake e as consequências sociais da mecanização da desconfiança

Lucia Santaella¹

Marcelo de Mattos Salgado²

Resumo: O artigo tem origem em trabalhos anteriores que analisaram a crise de confiança na sociedade e como esta seria paradoxalmente impulsionada, em parte, por avanços como o *blockchain*, tecnologia de segurança baseada em algoritmos de Inteligência Artificial. Ao mecanizar a confiança, o *blockchain* paradoxalmente reduz sua importância ou até torna a mesma desnecessária como elemento dos laços entre duas partes humanas. Tal inovação contribui para mudar gradualmente a dinâmica do sistema social da confiança interpessoal para uma sociedade do controle, ao substituir a confiança entre indivíduos por tecnologias de segurança e concentração de dados nas mãos de poucos – em particular, as big techs e burocracias estatais. Já a deepfake é uma manipulação audiovisual que também se baseia na contemporânea tecnologia algorítmica do *deep learning* e do *machine learning*. No entanto, o objetivo da deepfake é criar representações e percepções alternativas – e propositadamente falsas – da realidade, efetivamente mecanizando a desconfiança entre seres humanos. Este texto procura avaliar, desde o trabalho anterior, as consequências do deepfake para as relações humanas e a sociedade no contexto da crise de confiança.

Palavras-chave: Deepfake. Inteligência Artificial. Confiança. Controle. Redes digitais.

¹ Lucia Santaella é pesquisadora IA 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.

² Marcelo de Mattos Salgado é doutorando na PUC-SP (TIDD), jornalista e professor. Estuda o aumento da polarização nas redes digitais desde 2016. ORCID: orcid.org/0000-0001-8243-4977. CV Lattes: lattes.cnpq.br/9529682415224917. E-mail: msalgadosp@gmail.com.

Deepfake and the social consequences of the mechanization of distrust

Abstract: The article originates from previous works that analyzed the crisis of trust in society and how this would be driven, in part, by advances such as blockchain, a security technology based on Artificial Intelligence algorithms. By mechanizing trust, blockchain paradoxically reduces its importance or even makes it unnecessary as an element of the bonds between two human parts. Such innovation contributes to gradually change the dynamics of the social system from interpersonal trust to a society of control, replacing trust between individuals with security technologies and data concentration in the hands of a few – in particular, big techs and state bureaucracies. Deepfake is an audiovisual manipulation that is also based on the algorithmic technology of deep learning and machine learning. However, the purpose of deepfake is to create alternative – and purposefully false – representations and perceptions of reality, effectively mechanizing the distrust between human beings. This text seeks to assess, from the previous work, the consequences of deepfake for human relations and society in the context of the crisis of trust.

Keywords: Deepfake. Artificial Intelligence. Trust. Control. Digital networks.

Introdução

A crise de confiança, tanto de pessoas em instituições (governos, empresas, mídia e ONGs) quanto dos cidadãos entre si ainda vivida por países ocidentais, em geral, foi estabelecida e analisada em trabalho anterior (SALGADO, 2020). Este contexto será brevemente atualizado e estendido à China, sobretudo por conta da pandemia de Covid-19, com início em torno de fevereiro de 2020. A nova praga e as medidas adotadas contra ela continuam a arrasar o planeta de muitas formas – desde as mais evidentes milhões de mortes e questões de saúde física e mental, até questões políticas, econômicas e de controle social. Este quadro produz, também, mudanças severas na confiabilidade entre indivíduos e grupos mundo afora.

Segundo pesquisa da consultoria Edelman, cujo *trust index* (índice de confiança) calcula a média da confiabilidade dos cidadãos em governos, empresas, mídia e ONGs, a variação entre novembro de 2019 e novembro de 2020 na China foi de -10 pontos percentuais – de 82% para 72% – a maior queda do mundo, no período (EDELMAN, 2021, p. 9). O interessante é que, historicamente, a população chinesa tende a confiar (ou a responder que confia) muitíssimo em suas instituições, o que torna o registro desta queda abrupta ainda mais notável. Para uma comparação quanto ao referido índice de confiança em novembro de 2020: os EUA estavam com 48%; Brasil, 51%; Alemanha, 53%; e África do Sul, 48%. A mesma pesquisa indica a gravidade do impacto que a pandemia de Covid-19 teve sobre a fieza das pessoas a respeito de, particularmente, as burocracias governamentais (ibid., p. 5 e 11), com quedas importantes em países europeus e na China. Por outro lado, as empresas surgem como único ator – entre governos, empresas, mídia e ONGs – a reunir, ao mesmo tempo, competência e ética aos olhos das populações analisadas (ibid., p. 6-7).

O interregno da pandemia

A queda da fé na mídia e no jornalismo profissionais mundo afora, que já acontece há vários anos (EDELMAN, 2021; BRENAN, 2020; NÓBREGA, 2020; TWENGE, CAMPBELL; CARTER, 2014), viveu momentos de certa calma, com perdas menores e até ganhos de confiança, sobretudo na primeira metade de 2020 (NEWMAN, 2020, p. 10; NEWMAN, 2021, p. 18). Os possíveis motivos: audiência cativa de pessoas com medo, vulneráveis e presas em suas casas por conta de *lockdowns* e medidas similares, a partir da pandemia; e/ou seu retorno voluntário a fontes tradicionais. No entanto, os dados mais tardios de 2020 já trazem, de novo, quedas marcantes de confiabilidade das mídias profissionais (BRENAN, 2020; NOBREGA, 2020). A Edelman também registra esse desgaste continuado: de acordo com sua pesquisa mais recente, em médias globais até os primeiros meses de 2021, 59% dos entrevistados acreditam que “jornalistas e repórteres estão intencionalmente tentando enganar as pessoas dizendo coisas que eles sabem ser falsas ou exageros grosseiros”; 59% pensam que “a maioria das organizações noticiosas estão mais preocupadas com apoiar uma ideologia ou posição política do que informar o público”. Por fim, 53% das pessoas dizem confiar em mídias tradicionais em 2021, queda de 8% em relação a 2020, quando eram 61%; e um declínio de 12% em relação a 2019, quando 65% diziam confiar na imprensa (EDELMAN, 2021, p. 24 e 25). A Edelman considera “confiável” um resultado entre 60% e 100%.

Detalhe curioso do mesmo estudo da Edelman: a derrocada da credibilidade em informações obtidas nos mecanismos de pesquisa, tais como Google, responsável por 92,2% (STATCOUNTER, 2021) das buscas feitas em todo o mundo. Em 2017 e 2019, tais instrumentos teriam alcançado seu ponto mais alto de fé: 65%. Agora, em 2021, esta fé no Google e em sistemas afins despencou para 53%, sua maior queda histórica. Algo mais particular – em vez de informações, notícias – foi obtido pela Reuters em 2019 (NEWMAN, 2019, p. 21), que analisou a baixa confiança em notícias visualizadas a partir de mecanismos de busca (33%); e em redes digitais (23%). A exceção novamente foi, de forma ainda mais específica, a cobertura da pandemia de Covid-19, sobretudo na primeira parte de 2020, quando notícias obtidas em mecanismos de busca obtiveram 45% de confiabilidade e aquelas encontradas em redes digitais, 26% – números que, mesmo assim, são bastante modestos.

Ainda no sentido de compreender a situação de crescente descrença popular em dados obtidos em mecanismos de pesquisa e grandes redes digitais, vale mencionar o livro “Covid-19: o grande *reset*”, escrito pelo presidente do Fórum Econômico Mundial, Klaus Schwab, com Thierry

Malleret. O grande *reset* – grosso modo, mudanças sugeridas na gestão global, ou seja, de todos os países, em política, economia e meio ambiente – é proposto por Schwab como um caminho para o mundo seguir desde supostas lições com a pandemia de Covid-19 e em um ambiente internacional de queda generalizada da confiança no contrato social (SCHWAB; MALLERET, 2020, p. 73–74). Em certo momento, os autores refletem sobre a colaboração entre Apple e Google desde abril de 2020 para criar um aplicativo que permita a oficiais de saúde fazer a engenharia reversa dos passos e conexões de pessoas infectadas pelo vírus, rastreando suas vidas – o que provoca em parte considerável do público inquietação e até medo sobre crescente vigilância digital e controle estatal e empresarial (sobretudo, por parte das companhias *big tech*). A partir daí, passa a existir um problema básico que restringe a eficiência de aplicativos como o criado por Apple e Google:

A pessoa que carrega o aparelho móvel teria que voluntariamente baixar o aplicativo e concordar em compartilhar seus dados, e as duas companhias deixaram claro que sua tecnologia não seria fornecida a agências de saúde pública que não cumpram com suas diretrizes de privacidade. (SCHWAB; MALLERET, 2020, p. 124)

De outro modo: mesmo com a garantia das gigantes digitais – e dos governos que decidem adotar tais aplicativos – que regras de proteção à privacidade (criadas por aqueles atores) seriam seguidas, a adesão do público em geral é ainda muito baixa, o que prejudica a criação de uma base de dados e a pretensa eficiência em controlar a propagação do vírus. A principal razão por trás do receio de cidadãos comuns por todo o mundo em ceder seus dados, certa ou equivocada, é evidente. A rigor, é um dilema já clássico, mas ainda mais perceptível em sua versão atualizada: deve-se confiar em burocracias e empresas e renunciar à própria privacidade para, em teoria, estar mais seguro e protegido de um vírus perigoso?

Nos próximos meses e anos, o equilíbrio entre benefícios de saúde pública e perda de privacidade será avaliado cuidadosamente, tornando-se o assunto de muitas conversas animadas e debates acalorados. A maioria das pessoas, temerosa do perigo representado pela Covid-19, perguntará: não é tolice não aproveitar o poder da tecnologia para vir em nosso resgate quando somos vítimas de um surto e enfrentamos uma situação de vida ou morte? Eles então estarão dispostos a renunciar a muita privacidade e concordarão que, em tais circunstâncias, o poder público pode legitimamente anular os direitos individuais. Então, quando a crise passar, alguns podem perceber que seu país se transformou repentinamente em um lugar onde não desejam mais viver. (SCHWAB; MALLERET, 2020, p. 127)

Schwab e Malleret citam mesmo os alertas de Yuval Harari e Evgeny Morozov – este último está convencido dos riscos de que a pandemia de Covid-19 pode levar a uma distopia, “um futuro negro de um estado de vigilância tecno-totalitário” (ibid., p. 129). Já Harari acredita ser falsa a escolha entre privacidade e saúde: “Podemos e devemos ter privacidade e saúde. Podemos escolher proteger nossa saúde e interromper a epidemia de coronavírus, não instituindo regimes de vigilância totalitários, mas sim empoderando os cidadãos” (HARARI, 2020). O autor acredita que isto pode ser feito a partir dos exemplos de Taiwan, Coreia do Sul e Cingapura, que “se apoiaram muito mais em testagem extensa, relatos honestos e na cooperação voluntária de um público bem-informado” (ibid.). Caso contrário, Harari sinaliza para o risco, com precedentes históricos, de medidas de controle social durarem mais do que o período mais grave da pandemia.

Você poderia, é claro, argumentar em favor da vigilância biométrica como uma medida temporária tomada durante um estado de emergência. Ela acabaria uma vez que a emergência acabasse. Mas medidas temporárias têm o péssimo hábito de sobreviver às emergências, especialmente porque sempre há uma nova emergência espreitando no horizonte (HARARI, 2020).

Em outro documento de julho de 2020, o Fórum Econômico Mundial, em parceria com a McGill University, também aborda a ascensão da Internet dos Corpos (IoB), desdobramento da Internet das Coisas (IoT) que implica coletar nossos dados físicos por meio de dispositivos que podem ser implantados, engolidos ou simplesmente usados, gerando enormes quantidades de informações relacionadas à saúde. Algumas dessas soluções, como rastreadores de fitness, são uma extensão da Internet das Coisas, mas agora a Internet dos Corpos passa a ter controle sobre funções vitais do corpo convertendo-o em fonte geradora de dados pessoais. Com isso, levanta-se um conjunto específico de oportunidades e desafios, desde questões de privacidade até questões legais e éticas.

Segundo este *briefing*, “a Internet dos Corpos transformou o corpo humano em uma plataforma de tecnologia, visto que gera enormes quantidades de dados biométricos e comportamentais humanos” (WEF; MCGILL, 2020, p. 7). Essa penetração ainda mais íntima na personalidade e na privacidade, entretanto, não evoca confiabilidade por parte de muitos cidadãos comuns – e com algum fundamento.

Há uma consciência cada vez maior da vulnerabilidade de acessórios vestíveis [*wearables*] e dispositivos médicos ligados à Internet das Coisas a hackers e ataques cibernéticos, que expõem vidas humanas a potenciais danos físicos e riscos de privacidade. [...] Pesquisadores encontraram sérias falhas de segurança em *smartwatches* infantis, que os hackers podem usar para rastrear crianças, obter acesso a áudio e fazer ligações telefônicas para elas. A privacidade é um fator importante que afeta a confiança dos consumidores e a adoção de dispositivos ligados à Internet dos Corpos. (WEF; MCGILL, 2020, p. 10)

Some a este cenário de grave crise da confiança – com especial desconforto a respeito de dados obtidos em meios digitais e concedidos a governos e *big techs* – a inserção de tecnologias com base em algoritmos de Inteligência Artificial (IA). Esta realidade se faz cada vez mais vívida no dia a dia de bilhões de pessoas – seja em uma simples pesquisa no Google, a encomenda de uma refeição no aplicativo iFood ou a busca por um namorado no OkCupid – e acrescenta elementos ainda mais complexos à discussão sobre confiança. Em trabalhos anteriores (SANTAELLA, 2020; SALGADO, 2020), abordamos o blockchain, software de criptografia assimétrica sob a forma de cadeias de código (WILLIAMS, 2019, p. 9). Seu objetivo inicial era garantir segurança nas transações com a moeda digital Bitcoin – no entanto, o blockchain vem se espalhando como medida de segurança muito mais abrangente.

Conforme Salgado (2020), o modo como o blockchain se relaciona à confiança, resumidamente, é o seguinte: a tecnologia criptográfica permite, em uma troca social e/ou econômica qualquer, a substituição da confiabilidade – elemento historicamente constitutivo de laços humanos – por mecanismos de segurança. O blockchain seria, assim, *trustless* ou “aconfiável” (WILLIAMS, p. 15) precisamente porque consiste na mecanização da confiança, o que tornaria a mesma confiabilidade irrelevante em uma troca administrada via *blockchain*: a segurança criptográfica é o que importa.

O deepfake no contexto da crise de confiabilidade

A partir desta prévia análise acerca do contexto de crise generalizada da confiança, este texto presente busca considerar como a tecnologia algorítmica emergente do deepfake se insere no mesmo cenário. Paris e Donovan (2019, p. 5) indicam que os primeiros exemplos mais amplamente reconhecidos de tecnologia deepfake datam de novembro de 2017, quando um usuário do fórum Reddit enviou uma série de vídeos com os rostos de atrizes famosas, como Gal Gadot e Scarlet Johansson, enxertados nos corpos de outros atores em situações pornográficas:

Desde então, a mídia noticiosa e, portanto, o público em geral, começaram a usar o termo ‘*deepfake*’ para se referir a este gênero de vídeos que usam alguma forma de aprendizado ‘profundo’ ou de máquina para hibridizar ou gerar corpos e rostos humanos. (PARIS; DONOVAN, 2019, p. 5)

Em uma definição muito sucinta, o deepfake consiste, sobretudo, em manipulação audiovisual (ibid.). O deepfake é produto de aplicações de IA que fundem, combinam, substituem e sobrepõem imagens e clipes de vídeos para criar vídeos falsos, que parecem autênticos, de pessoas dizendo qualquer coisa em situações de caráter humorístico, pornográfico e/ou político – sem seu consentimento (WESTERLUND, 2019, p. 39). Isto significa que o deepfake, seja de forma intencional ou acidental, comumente distorce a percepção de terceiros a respeito de um indivíduo, associando seu nome e identidade a ideias e/ou atos que, por definição, não exprimem a realidade ou verdades sobre aquela pessoa. Karnouskos (2020, p. 1) considera os riscos e caos potencial, por exemplo, no simples cenário de uma teleconferência de uma empresa que não tem como avaliar precisamente quem está, de fato, presente: quais imagens são verdadeiras, de pessoas realmente conectadas, e quais são deepfakes.

O autor também traz um exemplo do que talvez possamos chamar de *metadeepfake*: o caso de um vídeo deepfake de 2018 com Barack Obama falando sobre os riscos dos deepfakes, algo que nunca acontecera. Por fim, Karnouskos registra a certa facilidade com que um usuário não avançado, com um computador doméstico e aplicativos disponíveis gratuitamente na Web, como o *DeepFaceLab* (ibid.), já pode criar deepfakes de níveis simples a intermediário – portanto, capazes de ludibriar os sentidos de muitos. Este fenômeno dos deepfakes feitos por indivíduos comuns, com tecnologia ordinária, é denominado por outros autores de *cheap fakes* (em tradução literal, “falsos baratos”). Os *cheap fakes* também são deepfakes, ou seja, tratam de manipulação audiovisual – apenas de forma menos sofisticada. Paris e Donovan caracterizam com clareza as diferenças entre deepfake e *cheap fake* e já apontam as complexas consequências para sociedade, cultura, política e, até mesmo, para o conceito de verdade – certamente próximo da confiança – que, neste contexto tecnológico, ganharia um caráter mais nitidamente relativista e socioconstrutivista.

O processo deepfake é o meio mais dependente do computador e também o menos publicamente acessível de criar mídia enganosa. Outras formas de manipulação audiovisual – “cheap fakes” – dependem de software barato e acessível, ou de nenhum software. Ambos deepfakes e cheap fakes são capazes de borrar a linha entre expressão e evidência. Ambos podem ser usados para influenciar a política das evidências: como as evidências mudam e são alteradas por sua existência em estruturas culturais, sociais e políticas. (PARIS; DONOVAN, 2019, p. 2-3)

A cobertura de notícias afirma que os deepfakes estão prestes a destruir a reivindicação da verdade por meio dos vídeos, borrando permanentemente a linha entre o vídeo comprobatório e o expressivo. Mas o que a cobertura desse fenômeno deepfake muitas vezes ignora é que a “verdade” do conteúdo audiovisual nunca foi estável – a verdade é social, política e culturalmente determinada. E as pessoas são capazes de manipular a verdade com deepfakes e cheap fakes. (Ibid., p. 6)

Por tudo que vimos até agora, não é difícil perceber a conexão multifacetada e profunda entre questões que giram em torno da confiabilidade e do deepfake. Por exemplo: distância entre público e privado, privacidade, anonimato e o que é ou não associado – inclusive falsamente – a nosso nome e identidade. Afinal, “[o] respeito está ligado aos nomes” e “[p]ode-se definir confiança como crença nos nomes” (HAN, 2018, p. 14–15). Becker (1996, p. 44) lembra que a confiança é um elemento das relações interpessoais que envolve nossas expectativas a respeito dos outros – por exemplo, de alguma ação (HAWLEY, 2014, p. 7) e/ou quanto a manter sua palavra. A confiabilidade trata também, pois, da esperança que depositamos em indivíduos específicos e de nossa percepção e crença na reciprocidade de atos, como sugere Adam Smith (1984, p. 337): “Nós confiamos no homem que parece disposto a confiar em nós”.

No entanto, como criar expectativas sobre os outros – e, em tempo, laços sólidos de fiesza – no atual contexto já estabelecido de crise de confiança, então salpicado por tecnologias, que, por um lado, substituem a fides por segurança criptográfica, caso do *blockchain*; e, por outro, recursos que permitem a criação de manipulações audiovisuais cada vez mais convincentes, como os deepfakes?

A crença de correspondência a fatos, que o deepfake cria e da qual é quase impossível se livrar, pode ser explicada pelo funcionamento do sistema perceptivo humano, a saber, o ser humano não está mentalmente equipado para duvidar daquilo que seus olhos veem. Isso não significa que a percepção não erra, mas sim, que, para corrigir o erro, um dado perceptivo precisa ser comparado a outro. É apenas em função dessa comparação que é possível constatar um erro ou equívoco perceptivo (SANTA-ELLA, 2012, p. 89-138).

Além disso, para complementar as consequências nefastas do deepfake, há o elemento complicador das redes digitais, que permitem uma propagação muitas vezes viral de seu conteúdo (KARNOUSKOS, 2020, p. 2). Entre outras questões, as redes digitais contribuem para extinguir a distância e a diferença entre o público e privado: “A falta de distância leva a que o privado e o público se misturem. A comunicação digital for-

nece essa exposição pornográfica da intimidade e da esfera privada. Também as redes sociais se mostram como espaços da exposição do privado” (HAN, 2014, p. 13). Uma vez que uma das tônicas do deepfake consiste em explorar situações ligadas à moralidade, especialmente em seus aspectos risíveis, é importante verificar o tipo de abrigo específico que as redes dão ao deepfake e em qual medida intensificam a crise de confiança generalizada que estamos vivendo.

A exibição do espaço privado nas redes

A desconfiança foi e é estudada por muitos autores em áreas como filosofia, psicologia e sociologia – nem por isso, ela faz por merecer uma atenção suficientemente distinta. Jason D’Cruz indicou a importância desta diferenciação: “A confiança é comumente descrita usando a metáfora de uma ‘cola social’ invisível que só chama a atenção quando está ausente. É inicialmente tentador pensar na desconfiança como a ausência da ‘cola’ da confiança. Mas essa maneira de pensar simplifica demais a relação conceitual entre confiança e desconfiança” (D’CRUZ, 2020, p. 41).

A partir de outras referências, o autor delinea que a desconfiança geralmente vem acompanhada de sentimento de insegurança, cinismo, desprezo e medo, ou seja, trata de uma postura ativamente negativa, em vez da certa neutralidade de um modo agnóstico do tipo “aguarde e observe”. É importante distinguir, ainda, a desconfiança de ordem moral da mera desconfiança normativa quanto a uma habilidade particular, por exemplo (ibid., p. 42). Desconfiar que uma pessoa é uma assassina é bem mais grave do que suspeitar de sua capacidade de cozinhar – ou de seu tato social. A desconfiança de teor moral é, certamente, sua versão crucial também em relação aos deepfakes e seus efeitos potencialmente mais devastadores entre pessoas, comunidades, confiança e tecido social.

Katherine Hawley, outra estudiosa sobre temas em torno de confiança e desconfiança, define ambas quanto a expectativas de que uma pessoa nos faça algo da seguinte forma: “Confiar em alguém para fazer algo é acreditar que ela tem o compromisso de fazê-lo e confiar nela para cumprir esse compromisso. Desconfiar de alguém para fazer algo é acreditar que ela tem o compromisso de fazê-lo, mas não confiar nela para cumprir esse compromisso” (HAWLEY, 2014, p. 10). A autora distingue “confiança” de “contar com algo ou alguém” (em inglês, *reliance*, frequentemente traduzido para a língua portuguesa como “confiança”). “Conto com a estante para que suporte o vaso [...] eu não confio na estante. Mas

também não desconfio dela” (ibid., p. 3). Deve-se notar que “contar com” também se aplica a pessoas: como diz Hawley, não contar com os outros não significa, necessariamente desconfiar deles. “Meus colegas nunca me compraram champanhe, então, em particular, não conto com eles para que me comprem champanhe na próxima sexta-feira” (ibid.).

No caso das deepfakes, as questões centrais parecem muito mais frequentemente – e mesmo, parcialmente, por definição – ocorrer a respeito de verdade ou mentira, com caráter pessoal e moral. Desta forma, a desconfiança – acompanhada comumente por insegurança, cinismo, desprezo e medo – seria uma consequência inescapável da manipulação audiovisual com objetivo de produzir conteúdo falso, algo que, a depender da qualidade do deepfake e da capacidade e atenção de seu intérprete, pode fabricar uma quantidade ilimitada de ruídos e leituras equivocadas sobre pessoas e, desta forma, sobre a própria realidade.

Finalmente, o efeito repetitivo das deepfakes nas redes e cumulativo ao longo de meses e anos somado a outros fatores discutidos em textos anteriores e à própria crise generalizada de confiança merece atenção. Se há algo bem estabelecido sobre a natureza humana é que nós aprendemos e adaptamos nosso comportamento: é bastante plausível propor que, se determinadas mídias, canais, grupos ou pessoas constantemente nos expõem a conteúdo falso, mentiroso ou distorcido – como deepfakes ou afins –, nosso grau de desconfiança em relação àqueles aumentará mais ainda. Ou seja, tanto confiança quanto desconfiança são modulares e podem variar em grau e magnitude ao longo do tempo, conforme as pesquisas relatadas no início deste artigo comprovam – não é uma questão binária, tal que há confiança ou não há; ocorre uma acumulação de evidências que empurra em uma ou na outra direção. Em outras palavras: quanto menor for a confiabilidade, por exemplo, de informações obtidas em redes digitais ou mecanismos de pesquisa *online*, mais grave fica a situação da estabilidade social e mais difícil será recuperar a confiança das pessoas.

Considerações finais

A partir deste artigo e de trabalhos anteriores, é possível destacar um dilema ou enigma que surge como consequência dos referidos estudos quanto ao uso de tecnologias digitais baseadas em algoritmos de IA – especificamente, os casos do *blockchain* e da deepfake – na conjuntura da grave crise de confiança mundial, com ênfase a ambientes digitais. Por

um lado, com o *blockchain*, temos a *mecanização da confiança* a partir da substituição de confiabilidade por segurança criptográfica, o que paradoxalmente elimina a confiança como laço humano historicamente experienciado da equação – ou seja, da troca social e/ou comercial.

Por outro lado, desde o sofisticado simulacro da deepfake, temos uma mais óbvia transgressão da verdade e o processo aqui denominado *mecanização da desconfiança*, que – diferente do *blockchain* – não traz consigo nenhum paradoxo ou complexidade. De outro modo: a deepfake realiza de fato o que se propõe a fazer e tende a propagar falsidades e informações erradas, muitas vezes desmoralizantes, a respeito de pessoas e grupos de tal modo a incitar reações equivocadas, ruídos comunicacionais e, no limite da lei, os crimes de difamação e calúnia. Desta forma, não é difícil concluir que, a exemplo do *blockchain*, mas de outra maneira, a deepfake também contribui, mesmo que de modo fragmentado e que, a rigor, deve ser analisado caso-a-caso, para o aprofundamento da crise de confiança que já abala o mundo.

O dilema e enigma, pois, dizem respeito exatamente à situação apresentada: se uma tecnologia de IA criada com a intenção de melhorar a segurança das trocas (*blockchain*) termina por substituir e eliminar a confiabilidade, reduzindo sua vivência como laço humano; e outra aplicação (deepfake) voltada para manipulações audiovisuais falsas efetivamente e ativamente aflige a confiança pessoal e social, o que pode ser feito? Haverá alguma tecnologia digital e, mais precisamente, algo baseado em algoritmos de IA que possa mitigar a severa crise de confiança generalizada que aflige o mundo – ou seria um padrão inescapável?

Referências

BECKER, Lawrence. Trust as noncognitive security about motives. *Ethics*, Chicago, v. 107, p. 43-61, 1996.

BRENAN, Megan. Americans remain distrustful of mass media. *Gallup. Politics*, September 30, 2020. Disponível em: news.gallup.com/poll/321116/americans-remain-distrustful-mass-media.aspx. 2020. Acesso em: 29 abr. 2021.

D'CRUZ, Jason. Trust and distrust. In: SIMON, Judith (ed.). *The Routledge handbook of trust and philosophy*. London: Routledge, 2020, p. 41-51.

EDELMAN Trust Barometer 2021. Disponível em: edelman.com/sites/g/files/aatuss191/files/2021-01/2021-edelman-trust-barometer.pdf. 2021. Acesso em: 29 abr. 2021.

HAN, Byung-Chul. *A sociedade da transparência*. Lisboa: Relógio D'Água, 2014.

_____. *No exame*. São Paulo: Vozes, 2018.

HARARI, Yuval. The world after coronavirus. *Financial Times*, March 20, 2020. Disponível em: [ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75](https://www.ft.com/content/19d90308-6858-11ea-a3c9-1fe6fedcca75). Acesso em: 29 jun. 2021.

HAWLEY, Katherine. Trust, distrust and commitment. *Noûs*, Hoboken, NJ, v. 48, n. 1, 2014, p. 1–20. Disponível em: onlinelibrary.wiley.com/doi/epdf/10.1111/nous.12000. 2014. Acesso em: 29 jun. 2021.

KARNOUSKOS, Stamatias. Artificial Intelligence in digital media: the era of deepfakes. In: *IEEE Transactions on Technology and Society*, v. 1, n. 3, Sept. 2020, p. 138-147. Disponível em: researchgate.net/publication/342795647_Artificial_Intelligence_in_Digital_Media_The_Era_of_Deepfakes. 2020. Acesso em: 28 jun. 2021.

LIU, Xiao et al. Shaping the future of the Internet of bodies: new challenges of technology governance (Briefing Paper, World Economic Forum, July 2020). Disponível em: weforum.org/docs/WEF_IoB_briefing_paper_2020.pdf. 2020. Acesso em: 10 jul. 2021.

NEWMAN, Nic, et al. *Reuters Institute Digital News Report 2019*. Oxford. Disponível em: reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf. 2019. Acesso em: 9 jun. 2021.

NEWMAN, Nic, et al. *Reuters Institute Digital News Report 2020*. Oxford. Disponível em: reutersinstitute.politics.ox.ac.uk/risj-review/google-and-university-oxford-agree-extension-support-digital-news-project-august-2020. Acesso em: 9 jun. 2021.

NEWMAN, Nic, et al. *Reuters Institute Digital News Report 2021*. Oxford. Disponível em: reutersinstitute.politics.ox.ac.uk/digital-news-report/2021. Acesso em: 12 jul. 2021.

NOBREGA, Ighor. PoderData mostra queda de confiança dos brasileiros na imprensa. *Poder 360*, 29 dez. 2020. Disponível em: poder360.com.br/poderdata/poderdata-mostra-queda-de-confianca-dos-brasileiros-na-imprensa/. 2020. Acesso em: 8 jun. 2021.

PARIS, Britt; DONOVAN, Joan. *Deepfakes and cheap fakes*. Thousand Oaks: Sage (=Data & Society's Media Manipulation research initiative). Disponível em: datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf. 2019. Acesso em: 29 abr. 2021.

- SALGADO, Marcelo. Blockchain e a crise de confiança na sociedade do controle. In: SANTAELLA, Lucia (org.). *A expansão social do blockchain*. São Paulo: EDUC, 2020. p. 25-39.
- SANTAELLA, Lucia. *Percepção: fenomenologia, ecologia, semiótica*. São Paulo: Cengage Learning, 2012.
- SANTAELLA, Lucia. Blockchain: de onde veio, onde está e para onde vai. In: SANTAELLA, Lucia (org.). *A expansão social do blockchain*. São Paulo: EDUC, 2020. p. 11-24.
- SCHWAB, Klaus; MALLERET, Thierry. *Covid-19: The great reset*. La Vergne, TN: Lightning Source, 2020.
- SMITH, Adam. *The theory of moral sentiments*. Indianapolis, IN: Liberty Fund, 1984.
- STATCOUNTER GlobalStats. Browser market share worldwide, July 2020-July 2021. Disponível em: gs.statcounter.com/search-engine-market-share. 2021. Acesso em: 8 jun. 2021.
- TWENGE, Jean; CAMPBELL, W. Keith; CARTER, Nathan. Declines in trust in others and confidence in institutions among American adults and late adolescents, 1972-2012. *Psychological Science*, v. 25, n.10, p. 1914-1923, 2014. Acesso em: 28 jun. 2021.
- WESTERLUND, Mika. The emergence of deepfake technology: a review. *Technology Innovation Management Review*, v. 9, n.11, p. 39-52, Nov. 2019. Acesso em: 29 abr. 2021. p. 39-52.
- WILLIAMS, Stephen. *Blockchain: the next everything*. New York, NY: Scribner, 2019.