

Com avanço tecnológico, fake news vão entrar em fase nova e preocupante

Segundo autores, é preciso discutir o que deu errado com a internet e repensar o contrato social da informação

*** Virgílio Almeida**

Danilo Doneda

Ronaldo Lemos

O vídeo chama a atenção. Michelle Obama, ex-primeira dama dos Estados Unidos, veste uma blusa decotada, sorri e começa a se despir para a câmera. Parece real, mas não é; trata-se de exemplo de "deep fake" (falsificação profunda), produzida com auxílio de inteligência artificial.

O leigo talvez demore a aceitar que aquelas imagens não são verdadeiras e não foram filmadas, mas sintetizadas digitalmente. Essa tecnologia, antes acessível somente para Hollywood, está cada vez mais à disposição.

Pode-se aplicar o rosto de alguém em cenas de sexo ou em qualquer situação comprometedora. O mesmo vale para a voz. Com amostras da fala de uma pessoa, um software faz com que ela diga qualquer coisa, com timbre, cadência e entonação próximos da perfeição.

Ou seja, a divulgação de textos com conteúdo de algum modo mentiroso representa apenas o começo do fenômeno das *fake news*. O próximo passo parece ser a era das *deep fake news*. Será cada vez mais difícil separar a realidade da manipulação digital — e há quem diga que estamos vivendo os últimos dias da própria ideia de realidade, como fazem Claire Wardle e Hossein Derakhshan em recente relatório do Conselho da Europa.

Uma outra dupla de autores (Danielle Citron, da Universidade de Maryland, e Robert Chesney, da Universidade do Texas) publicou em fevereiro um artigo na revista *Lawfare* no qual discutem como as "deep fakes" tornam a manipulação ainda mais perniciosa.

Diante de vídeos ou gravações que parecem reais, qual reputação resistiria a um ataque? Como saber que é falsa a cena em que um político diz coisas odiosas ou recebe propina?

Essa não é uma ameaça apenas para indivíduos, mas para os próprios pilares do convívio democrático. Um vídeo falso pode ser usado para provocar convulsão social. Por exemplo, ao mostrar um policial atirando num civil inocente, ou um bandido assaltando transeuntes numa dada localidade.

Antes de haver tempo para denunciar a falsificação profunda, as imagens talvez tenham suscitado reações bem reais, como o linchamento de um inocente.

Dado o perigo crescente que as fake news representam, não surpreende que cada vez mais atores se mobilizem para combatê-las. O fenômeno em si, contudo, não é novo.

Há registros de manipulação de informação com o objetivo de influenciar o processo político desde ao menos o Império Romano. Num passado menos distante, o jornal New York Sun fez sucesso em 1835 com a publicação de artigos alardeando a descoberta de vida na Lua. Mais recentemente, em 2006, uma rede pública de TV na Bélgica afirmou que o Parlamento flamengo havia declarado a independência da região.

CLASSIFICAÇÕES

Os fatores por trás da desinformação são diversos. Vão da simples negligência (como a disseminação de boatos ou matérias jornalísticas mal-apuradas) à busca de vantagens políticas ou financeiras, passando pela tentativa de destruir reputações.

Levando em conta essas gradações, o Conselho da Europa classificou as fake news em três categorias dentro um quadro maior que chamou de desordem informacional, um conceito relevante por abranger diversas nuances da manipulação.

Uma delas é a desinformação ("disinformation"), que consiste em notícias falsas deliberadamente criadas e espalhadas para prejudicar uma pessoa, um grupo social, uma organização ou um país.

Outra é a notícia falsa propriamente dita ("misinformation"), compartilhada por uma pessoa desavisada que a princípio não tinha a intenção de prejudicar alguém. Como aqui o critério não é a má-fé, incluem-se até reportagens com erros causados por falhas na apuração.

E, por fim, o que chamou de "mal-information" (malinformação), notícias que, embora tenham bases reais, são editadas e disseminadas com a finalidade de causar danos — por exemplo, revelando publicamente temas da esfera privada.

Essas categorias poderiam descrever o fenômeno das fake news em qualquer período da história, mas dois elementos fundamentais são específicos da atualidade: a velocidade com que as notícias falsas se espalham e a capilaridade que elas têm.

Hoje, plataformas como Facebook, Twitter, Google e YouTube, entre outras, possuem alcance global instantâneo (ao menos no Ocidente).

A combinação dessa arquitetura com embates políticos nacionais, disputas geopolíticas globais e modelos de negócio baseados em publicidade comportamental criou as condições para que se aprofundasse o fenômeno da desordem informacional.

Vários países vêm desenvolvendo estratégias para enfrentar esse desafio contemporâneo. As dificuldades são inúmeras: além da necessidade de delimitar o problema (como definir o que é falso?), é preciso lidar com as diferentes características da rede. Cada nação consegue agir somente em relação ao seu próprio território.

Muitas soluções propostas conflitam diretamente com direitos fundamentais, como a liberdade de expressão e de informação ou o devido processo legal.

Algumas postulam mecanismos de intervenção na rede que não raro resvalam na censura. Outras propõem tratar notícias falsas geradas em outros países como tema de segurança nacional, o que autorizaria retaliação militar. Mesmo medidas extremas, contudo, podem se revelar inócuas.

É possível dividir as propostas em pelo menos dois grupos: as soluções de "hard power", com as quais governos miram os mecanismos de circulação da informação, e as de "soft power", com estratégias de longo prazo capazes de vacinar a sociedade

contra esse tipo de manipulação, em cooperação com diversos setores (universidades, governos, empresas privadas, sociedade civil etc.).

Em qualquer das hipóteses, a batalha contra a desordem informacional será travada sobretudo em dois campos distintos: ciência e tecnologia de um lado, direito e regulação de outro.

PESQUISAS

A ciência vem tendo papel importante. Estudos de Deb Roy (professor do Massachusetts Institute of Technology, o MIT) e colegas, publicados na revista *Science*, comprovam, por exemplo, que as notícias falsas circulam muito mais depressa e de maneira mais abrangente do que as verdadeiras.

Eles também constataram que a mentira possui um estilo de redação distinto da verdade —a mentira usa linguagem simples e direta e apela para sentimentos básicos, como medo e raiva.

Reações às notícias falsas tendem a expressar desgosto, temor e surpresa. Por sua vez, notícias verdadeiras costumam gerar comentários de apreensão, tristeza, alegria ou confiança.

A partir desses padrões, o grupo do MIT criou um algoritmo capaz de detectar com 75% de acerto se um rumor no Twitter é verdadeiro ou falso. Para isso, são examinados apenas o estilo linguístico do tuíte e a característica das reações a ele, bem como sua dinâmica de propagação —não há análise do conteúdo da mensagem.

Esse tipo de pesquisa abre caminhos novos e promissores para as atividades de checagem de fatos ("fact-checking"). Uma das grandes dificuldades nessa área é o volume imenso de informação a ser verificada. Se os algoritmos puderem auxiliar nesse trabalho, baseando-se inclusive na ciência dos dados, esse fardo pode se tornar mais leve, com efeitos positivos.

Mas por que as notícias falsas e outras formas de desinformação se disseminam tão rapidamente? Em outras palavras, por que a mentira circula mais e mais depressa que a verdade?

Dois componentes são centrais: de um lado, as pessoas que fazem parte do ciclo de criação e difusão da informação; de outro, as características das plataformas que atuam por meio de algoritmos e processos automáticos de decisão.

O papel das pessoas é referenciado no mesmo estudo de Roy. O grupo do MIT analisou histórias verdadeiras e falsas que circularam pelo Twitter desde o seu início, em 2006, até 2017. Foram analisadas 126 mil sequências de tuítes com mensagens em inglês, espalhadas mais de 4,5 milhões de vezes por 3 milhões de pessoas.

Em todos os temas observados (política, negócios, entretenimento, ciência, desastres naturais etc.), verificou-se que notícias falsas se espalham mais depressa, chegam mais longe e atingem mais pessoas do que as verdadeiras.

A probabilidade de notícias falsas serem retransmitidas é 70% maior do que a das verdadeiras; para temas políticos, essa percentagem aumenta.

Por ora, o estudo desmente a hipótese de que as pessoas que espalham notícias falsas sejam mais conectadas em redes sociais (com mais seguidores) ou usem melhor as funções das plataformas.

Após testes estatísticos, os pesquisadores demonstraram que o fator novidade é o elemento mais característico na diferença da disseminação de verdades ou mentiras.

Usando métodos computacionais e análises lexicográficas para inferir o contexto emocional nas comunicações, os pesquisadores descobriram que notícias falsas criam um senso de urgência e novidade. Isso atrai a atenção e incentiva o compartilhamento. Muitas notícias falsas são resultantes de um design minucioso, feito para inocular esse senso de urgência e novidade e, assim, incentivar sua disseminação. Com os novos estudos, começa a ficar claro que esse procedimento pode deixar uma impressão digital, o que facilitaria identificar e combater a desinformação.

Há ainda o fator da polarização política da mensagem. Em artigo recente no jornal *The New York Times*, a professora Zeynep Tufekci, da Universidade da Carolina do Norte, descreveu como se dá essa dinâmica no YouTube.

A ligação entre inteligência artificial e o modelo de negócio do Google/YouTube leva à busca permanente e incessante da atenção dos usuários —quanto mais tempo as pessoas permanecem na plataforma, maior o potencial de receita publicitária. Daí por que os algoritmos sugerem novos vídeos e os exibem automaticamente.

Tufekci descreve vários testes que podem ser repetidos por qualquer pessoa. Por exemplo, em buscas por conteúdo informativo normal, plataformas como o YouTube sugerem nas barras laterais vídeos cada vez mais radicalizados e conspiratórios. Ou seja, o usuário termina exposto a esse tipo de produto mesmo que não esteja procurando por ele.

O YouTube vem sendo questionado também em relação à ação voltada para crianças. Há alguns meses, o ensaio "Something is wrong on the internet" (algo está errado na internet), de James Bridle, teve grande repercussão ao descrever uma série de situações em que vídeos infantis são acompanhados de recomendações de conteúdo impróprio para menores.

Grupos que se interessam pela promoção da desordem informacional já perceberam essas propriedades dos algoritmos nas plataformas. Também perceberam que é possível direcionar conteúdo para públicos específicos. A coleta de dados de usuários da internet permite determinar com precisão as suscetibilidades a vários tipos de mensagem.

DADOS PESSOAIS

Tome-se o exemplo da empresa Cambridge Analytica, cujas práticas foram descortinadas pelo jornal britânico *The Guardian*.

Valendo-se de um estudo de psicométrica feito pelo pesquisador Michal Kosinski, da Universidade de Cambridge (que nada tem a ver com o caso e desde o início não quis que seu trabalho fosse utilizado para fins privados), a empresa criou um psicográfico em que classifica eleitores em vários perfis, que vão de extrovertido a neurótico.

A campanha eleitoral torna-se, assim, menos uma troca de ideias do que uma operação de manipulação psicológica.

Como esse tipo de iniciativa depende de informações pessoais de boa qualidade e em abundância, o controle efetivo desses dados por parte do usuário é fundamental no combate às fake news.

Ao que tudo indica, nas recentes eleições da França, da Alemanha e da Itália, a desordem informacional foi quantitativamente menor e menos eficaz porque esses países contam com regras que dificultam a coleta de dados pessoais.

Medidas restritivas nesse campo passarão a valer em toda a União Europeia a partir de maio, quando entrará em vigor o seu Regulamento Geral de Proteção de Dados. No

Brasil, debate-se há anos a promulgação de uma Lei Geral sobre Proteção de Dados Pessoais; os projetos¹ estão na Câmara e no Senado.

Conforme notou James Ball em artigo sobre o Facebook no jornal britânico The Guardian, se dados pessoais são o novo petróleo, os seus poços estão cada vez mais concentrados. É crucial criar um modelo de negócio viável baseado na privacidade.

Outro aspecto importante a ser levado em conta é a descentralização da difusão de conteúdos na internet. Quando alguém repassa uma notícia, está atuando não só como destinatário mas também como meio para sua propagação.

Trabalho recente de um grupo de pesquisa da UFMG² mostrou a relação direta entre polarização política e desinformação. Usando tuítes referentes à disputa eleitoral dos EUA, o estudo mostrou que, quanto mais extremista for o usuário, mais associado à disseminação de fake news ele estará.

Esse fator ganha relevância devido à ascensão de ferramentas de comunicação pessoal como meio de divulgação de notícias falsas. Aplicativos como Telegram, Signal e WhatsApp, entre outros, não são mediados por algoritmos. Ou seja, nesses casos, prevalece a natureza humana, estimulada pelo design dessas plataformas e das redes que se formam por meio delas.

Quando a distribuição de conteúdo ocorre por meio de algoritmos, é possível disputar os parâmetros das fórmulas matemáticas para restringir a circulação de conteúdos radicalizados ou conspiratórios, por exemplo. No entanto, como fazer o mesmo em relação a plataformas de comunicação pessoal?

Há ainda outro elemento complicador. São os chamados bots (robôs), contas automáticas que imitam o comportamento humano.

Em muitas situações, eles são projetados para passar a impressão de que determinada postagem chamou a atenção de um número grande de usuários, o que leva os algoritmos das plataformas a reforçar a distribuição daquele conteúdo. Com isso, os bots ampliam o alcance das fake news e exploram as vulnerabilidades psicológicas ou sociais dos cidadãos.

Nesses casos, distorcem o debate público para atender a interesses políticos ou econômicos. Segundo artigo recente da Science, há de 29 milhões a 49 milhões de robôs no Twitter (de 9% a 15% das 330 milhões de contas ativas); no Facebook, são cerca de 60 milhões de bots controlando perfis (ou 3% dos mais de 2 bilhões de usuários).

Vale lembrar que um robô não é bom ou mau por si só. Essa ferramenta vem sendo utilizada para atendimentos online e para fiscalizar gastos públicos, entre vários exemplos de atuação positiva.

No entanto, todos esses robôs deixam claro que são robôs. Essa mesma atitude não existe em casos de manipulação política. Por isso, é urgente implementar regras para contas automatizadas, respeitando seis princípios: transparência, responsabilidade, inteligibilidade, acurácia, capacidade de auditar e equidade.

O QUE FAZER

A desordem informacional veio para ficar e vai se tornar cada vez mais complexa. Tecnologias de inteligência artificial usadas para criar bots, "deep fakes" e outras maneiras de mimetizar o comportamento humano estão ficando mais baratas, mais fáceis de usar e mais potentes.

Para piorar, em entrevista à revista Nature, o professor Hany Farid, da Universidade de Dartmouth, especialista no desenvolvimento de técnicas para detectar vídeos falsos, afirma que o campo de perícia forense digital está atrasado em relação ao vídeo.

Na tentativa de combater as fake news e outras formas de desinformação, nos últimos cinco anos surgiram inúmeras agências de verificação de fatos. Embora esses serviços tenham imensa relevância, não se sabe se as pessoas os utilizam em escala considerável ou se confiam neles.

De qualquer forma, sozinhas, as agências de "fact-checking" não resolvem as questões que emergem em ambientes altamente polarizados, quando a credibilidade da informação é questionada sempre que não esteja alinhada a determinada crença ideológica.

Por sua vez, soluções legislativas que procuram entregar para o Judiciário a tarefa de punir notícias falsas, arbitrando o que é falso ou verdadeiro, têm como consequência atacar diretamente a liberdade de expressão e de informação —e correm o risco de serem ineficazes.

O caminho que nos parece mais promissor é a abordagem multidimensional que foi proposta em recente relatório de um grupo de especialistas apontados pela Comissão Europeia. Essa proposta enfatiza as seguintes linhas de atuação:

- i) obter maior transparência na divulgação de notícias online, inclusive em relação a como os dados pessoais são usados para direcionar informações aos leitores;
- ii) promover competências em educação para a mídia ("media literacy"), a fim de auxiliar usuários a navegar num mundo com superabundância de informação;
- iii) desenvolver instrumentos e ferramentas para que jornalistas, aliados a cidadãos, possam combater a desinformação;
- iv) impulsionar a diversidade e a sustentabilidade dos meios de comunicação;
- v) estimular estudos continuados sobre o impacto da desordem informacional, tratando deles com análises científicas.

O fenômeno das fake news, em síntese, é uma oportunidade de discutir o que deu errado com a internet. Nos anos 1990 e 2000, a rede era vista como força de democratização e melhoria das condições planetárias, mas nesta década começa a prevalecer a percepção dos aspectos distópicos.

Sem perceber, estamos neste momento discutindo um novo equilíbrio para o ecossistema da informação. Para isso, é necessário um debate público constante e a participação ativa e consequente de todos os atores da sociedade.

A única forma de combater a pervasividade da desordem informacional será pela construção de um novo contrato social para a informação.

Virgílio Almeida, 67, é professor associado da Universidade Harvard. Membro titular da Academia Nacional de Engenharia (ANE). Foi professor titular do Departamento de Ciência da Computação da UFMG.

Danilo Doneda, 47, doutor em direito civil pela Uerj, é professor no Instituto de Direito Público (IDP) e especialista em privacidade e proteção de dados.

Ronaldo Lemos, 42, é advogado, diretor do Instituto de Tecnologia e Sociedade do Rio (ITS Rio), professor da Universidade de Columbia e colunista da **Folha**.

Fonte: publicado no jornal Folha de São Paulo em 08/04/2018