

Desinformação e as Eleições de 2020:

Como as Redes Sociais Deveriam se Preparar

PAUL M. BARRETT



Conteúdo

Sumário Executivo	1
1. Introdução	3
2. A Que Devemos Ficar Atentos em 2020	5
3. Qual a Resposta Dada por Empresas de Mídia Social a Ações de Desinformação	13
4. Conclusão e Recomendações	17
Notas de Fim	21

Agradecimentos

Este relatório se valeu das ideias de muitas pessoas, incluindo Kevin Carroll, da Wiggin and Dana; Joan Donovan, do Projeto de Pesquisa em Tecnologia e Mudança Social da Harvard Kennedy School; Yaël Eisenstat, do Centro de Tecnologia Humana; Lee Foster, da FireEye; Dipayan Ghosh, do Projeto Platform Accountability da Harvard Kennedy School; Brittan Heller, do Centro Carr de Política de Direitos Humanos da Harvard Kennedy School; Justin Hendrix, do NYC Media Lab; Jeff Jarvis, da Escola de Pós-Graduação em Jornalismo Craig Newmark da City University of New York; Darren Linvill e Patrick Warren, da Clemson University; Filippo Menczer, da Indiana University; Lisa-Maria Neudert, do Instituto para a Internet de Oxford; Nick Pickles, do Twitter; Iain Robertson e Andy Carvin, do Laboratório de Pesquisa sobre Crimes Digitais do Atlantic Council; Laura Rosenberger e Bret Schafer, da Aliança para a Garantia da Democracia; Alexandria Walden e Clement Wolf, do Google; e Candid Wueest, da Symantec

Agradecemos pelo apoio financeiro das Organizações Filantrópicas Craig Newmark, Fundações Open Society, Fundação Jonah Goodhart e John S. e James L. Knight

Autor

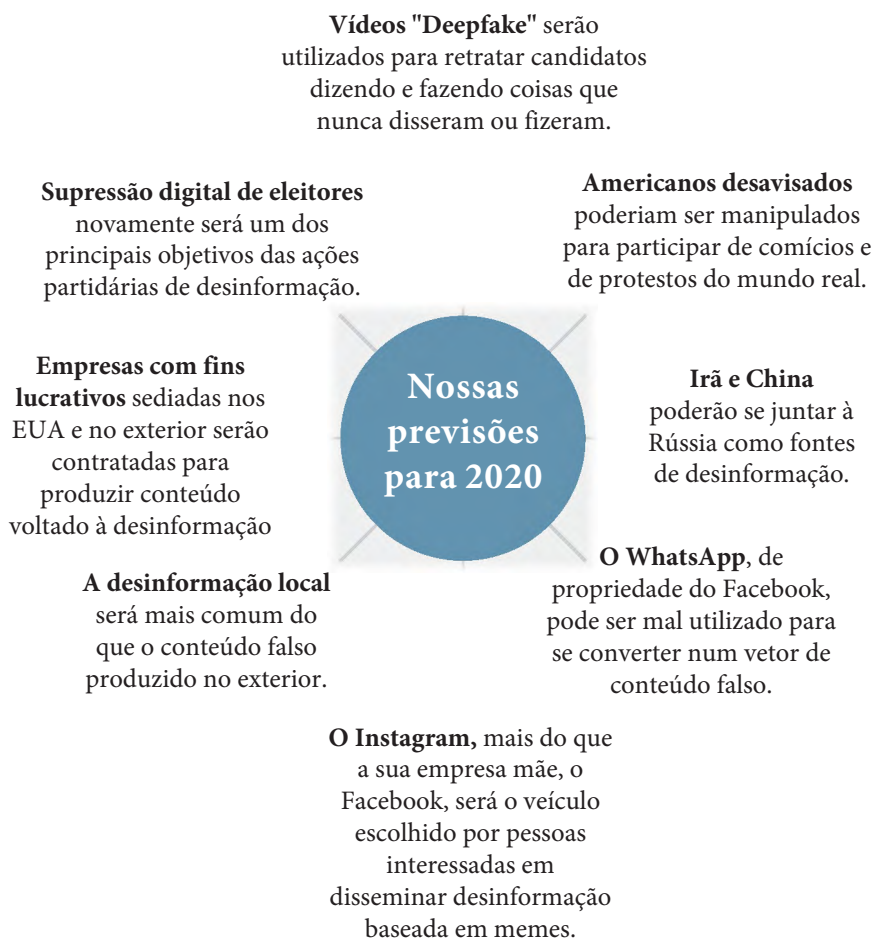
Paul M. Barrett é o diretor-adjunto do Centro Stern para Negócios e Direitos Humanos da New York University.

Sumário Executivo

“
Falando apenas de volume,
a desinformação produzida
no próprio país já
ultrapassa o conteúdo
nocivo oriundo de fontes
estrangeiras, certamente se
convertendo num fator nas
próximas eleições

”

Este relatório avalia algumas das formas e fontes de desinformação com potencial para alcançar destaque durante a campanha eleitoral presidencial de 2020. Muito embora não sido identificada muita interferência russa nas eleições minoritárias de novembro de 2018, não há qualquer garantia de que a Rússia, e outros antagonistas dos Estados Unidos, deixarão de interferir digitalmente nas eleições majoritárias de 2020. Além disso, falando apenas de volume, a desinformação produzida no próprio país já ultrapassa o conteúdo nocivo oriundo de fontes estrangeiras, certamente se convertendo num fator para as próximas eleições



O relatório explora esses riscos e analisa o que as principais empresas de mídia social —Facebook, Twitter e YouTube (de propriedade do Google) — fizeram para fortalecer suas defesas contra a desinformação. Também trazemos recomendações de medidas adicionais que as empresas deveriam adotar para enfrentar o ataque quase certo em 2020.

Terminologia e Escopo

Entende-se que "desinformação" seja uma categoria relativamente ampla de textos ou imagens falsos ou enganosos que se espalham de forma intencional ou por descuido, para enganar, estimular e promover a desavença, ou para arrecadar recursos por meio de esquemas do tipo "clickbait". Apenas para não utilizar o mesmo termo sempre, também fazemos uso dos termos "conteúdo falso" ou "informação falsa".

Também tratamos de uma categoria mais restrita de desinformação —"conteúdo comprovadamente falso"—, conteúdo que a nosso critério deveria ser eliminado pelas empresas de mídia social. Para ilustrar, vamos analisar estes exemplos hipotéticos: É possível analisar um artigo com o título "O massacre de Sandy Hook foi uma encenação" e chegar à conclusão, de forma inequívoca, de que se trata de conteúdo falso, e que por isso deve ser apagado. Por outro lado, uma história intitulada "Os jornalistas são os verdadeiros inimigos do povo" pode ter pouco fundamento, podendo até ser enganosa, mas não é possível comprovar que ela é falsa.

Este relatório não trata de ataques cibernéticos —invasões remotas de redes de computadores, como a investida russa de hackear e liberar a conta de email dos Democratas em 2016. Os ataques cibernéticos a campanhas e a sistemas de votação representam um outro perigo para 2020, mas estão fora do escopo deste relatório. Também estão fora do escopo as questões de privacidade e divulgação que levaram o Facebook a concordar em julho de 2019 em pagar US\$ 5,1 bilhões em acordos com o governo federal e a aceitar novas formas de supervisão.

Resumo das nossas Recomendações para as Empresas de Mídia Social

- 1 Detecte e remova vídeos deepfake.**
Vídeos realistas, mas fraudulentos, podem minar candidatos políticos e agravar o cinismo dos eleitores.
- 2 De forma geral, elimine o conteúdo comprovadamente falso.**
As plataformas já removem conteúdos classificados como discursos de ódio, supressão de eleitores e outras categorias. Recomendamos que acrescentem mais uma categoria.
- 3 Contrate um supervisor de conteúdo sênior**
Each company needs an executive with clout to supervise the process of guarding against disinformation.
- 4 Atenção ao problema de desinformação no Instagram.**
A plataforma de postagem de fotos requer atenção especial da sua empresa mãe, Facebook.
- 5 Limite o alcance do WhatsApp.**
Os usuários devem se restringir a encaminhar o conteúdo para um grupo de bate-papo por vez.
- 6 Combata ações de desinformação com fins lucrativos.**
As empresas devem estar preparadas para o conteúdo falso gerado por empresas contratadas para essa finalidade.
- 7 Apoie a legislação para regular anúncios políticos e punir a supressão de eleitores.** Projetos de lei bem específicos que estão pendentes no Congresso podem ajudar a conter algumas formas de desinformação.
- 8 Melhore o nível de colaboração do setor como um todo no combate à desinformação.** Por exemplo, quando uma plataforma derruba contas abusivas, outras devem fazer o mesmo com contas afiliadas.
- 9 Promova a educação em redes sociais de maneira mais direta e sustentada.** Os usuários precisam assumir a responsabilidade por reconhecer conteúdo falso, mas eles precisam de mais ajuda para fazer isso.

1. Introdução

“
O episódio envolvendo a congressista Nancy Pelosi prenunciou um tipo de desinformação que poderia atrapalhar a eleição de 2020: vídeo deliberadamente distorcido, amplificado pelas redes sociais
”

No fim de maio de 2019, dois vídeos pejorativos se espalharam pelas redes sociais. Eles retratavam a congressista Nancy Pelosi, presidente da Câmara dos Deputados dos Estados Unidos. Um desses vídeos, publicado numa página conservadora do Facebook chamada Politics WatchDog, divulgou em câmera lenta um discurso da congressista durante uma aparição pública, fazendo parecer que ela estaria embriagada. Rudolph Giuliani, um dos advogados pessoais do presidente Donald Trump, compartilhou o vídeo no Twitter, perguntando: "O que ocorreu com a Nancy Pelosi?" (Posteriormente, Rudolph Giuliani apagou o tweet).

Mais ou menos na mesma época, o próprio Trump tuitou um outro vídeo da Nancy Pelosi, agora editado pela Fox Business Network. A Fox fez cortes no vídeo de uma coletiva de imprensa, fazendo parecer que a Presidente estava gaguejando. "PELOSI GAGUEJA EM ENTREVISTA", declarou o presidente.¹

O episódio da Nancy Pelosi prenunciou um tipo de desinformação que provavelmente afetará as eleições de 2020: vídeos distorcidos de propósito, com efeito amplificado pelas redes sociais. A ameaça mais assustadora vem dos deepfakes, que são vídeos sintéticos criados com inteligência artificial (IA). Os personagens dos deepfakes parecem dizer ou fazer coisas que nunca disseram ou fizeram. Os vídeos da Nancy Pelosi, distorcidos de maneira grosseira, foram feitos com uma tecnologia disponível há muito tempo num formato conhecido como "cheapfakes". Demonstram que não é necessário ter conhecimentos avançados de informática para chamar a atenção

em massa para a representação enganosa de um inimigo famoso. Só o vídeo do discurso "bêbado" recebeu mais de três milhões de visualizações em alguns dias. Os deepfakes, que são mais difíceis de refutar, podem alcançar uma circulação ainda mais ampla.

Os vídeos da Nancy Pelosi apontam para outro tema deste relatório: No caso dos Estados Unidos, embora os operadores russos e outros atores estrangeiros quase sempre venham à tona (ou ressurgam) em 2020, um maior volume de desinformação provavelmente virá de fontes locais. Algumas dessas fontes locais serão sites obscuros ou contas de redes sociais, como o Politics WatchDog; outros serão mais conhecidos. Se ele seguir à sua linha, o Presidente Trump usará sua conta no Twitter (que hoje tem 62 milhões de seguidores e não para de crescer) para emitir declarações não verdadeiras. Até o início de agosto, ele já fez mais de 12.000 alegações falsas ou enganosas durante o seu mandato de presidente, muitas delas via Twitter, de acordo com o Fact Checker do Washington Post.²

Deepfakes e desinformação doméstica são duas das nossas oito previsões para 2020. Segue uma lista completa com o resumo das nossas previsões:

- **Deepfakes** ameaçam ofuscar a realidade num momento em que a existência de fatos objetivos é cada vez mais questionada.
- **Desinformação doméstica** virá principalmente da direita política, mas a esquerda também se entregará à distorção da verdade nas redes sociais.
- **Americanos desavisados:** a Agência de Pesquisa na Internet ou outras organizações russas poderão tentar recrutar americanos reais para atividades da vida real.
- O **Instagram**, serviço de imagem e vídeo de propriedade do Facebook, pode se tornar mais um veículo de desinformação ainda maior do que o Facebook, Twitter ou YouTube.
- O **WhatsApp**, serviço de mensagens do Facebook, pode se tornar uma fonte preocupante de conteúdo falso nos EUA, como aconteceu durante as eleições no Brasil e na Índia.
- **Io Irã e a China** são fontes potenciais de interferência estrangeira, mesmo quando a Rússia continua tentando irritar o eleitorado americano.
- Empresas com fins lucrativos com sede nos EUA e no exterior oferecerão seus serviços de manipulação de eleições.
- A **supressão digital de eleitores**, um meio altamente direto de afetar os resultados das eleições, mais uma vez representará um desafio para as redes sociais.

É impossível dizer com certeza como essas estratégias serão utilizadas em conjunto, e se outras também entrarão em cena. Parece seguro esperar que a interferência bem documentada da Rússia em 2016 —uma invasão com diferentes frentes, incluindo pessoas disfarçadas nas redes sociais, exércitos de contas automatizadas por robôs, e distribuição de e-mails hackeados dos democratas— não ocorrerá do mesmo jeito no próximo ano. Como afirmou o analista de desinformação Clint Watts: "Os truques do último ciclo não funcionarão da mesma forma em 2020".³

Políticas Confusas, Responsabilização Arbitrária

Além do prognóstico, este relatório examina a forma em que as empresas de mídia social reagiram às ações de desinformação com finalidade política. No caso do vídeo de "bêbada" da Nancy Pelosi, houve uma diferença notável entre as plataformas. O YouTube removeu prontamente as cópias do vídeo alterado. O Facebook, de propósito, não. Em vez disso, o Facebook classificou o vídeo como falso e reduziu a sua distribuição aos usuários. O Twitter também se recusou a eliminar o vídeo. É difícil generalizar a partir desse resultado, exceto para dizer que as políticas das plataformas são confusas e geralmente parecem ser aplicadas de maneira arbitrária.

Mantemos como princípio geral que, quando um conteúdo comprovadamente falso, como o vídeo da Nancy Pelosi, chama a atenção das redes sociais, ele deve ser retirado, e não apenas ser classificado e rebaixado. A prioridade deve ser dada à remoção de conteúdo comprovadamente falso que afeta a política ou as instituições democráticas. Proteger-se contra essa abundância de desinformação é uma tarefa difícil. Quinhentas horas de vídeo são enviadas para o YouTube a cada minuto. 4 Dado esse volume e a velocidade com que circula o conteúdo on-line, não seria razoável supor que toda desinformação possa ser removida. Mas isso não significa que as empresas não devam fazer nada além do que já estão fazendo. Elas estão numa posição privilegiada para monitorar as suas páginas com pessoas e com algoritmos, e acreditamos que elas são responsáveis por mitigar o dano que as ações de desinformação causam à nossa vida pública.

Para manter um registro de conteúdo comprovadamente falso, as plataformas deveriam manter uma base pesquisável isolada de conteúdo removido, por representar desinformação. Dessa forma, pesquisadores, jornalistas e usuários como um todo poderiam visualizar o conteúdo falso para pesquisas ou outros fins sem aparecer em resultados de pesquisa, em recomendações ou feeds de usuários comuns. Essa base de dados seria um registro do conteúdo digital mentiroso, sem contribuir

para a sua amplificação. E a base de dados serviria de referência para recursos dos usuários que acreditam que o seu post ou tweet foi removido inadequadamente.

Foco nas Empresas

Este é o quarto relatório do Centro Stern de Negócios e Direitos Humanos da NYU sobre desinformação online. Nossa publicação mais recente foi "**O Combate à Desinformação Local: O que as Empresas de Mídia Social Devem Fazer**" (março de 2019).⁵ Antes disso, publicamos "**O Combate à Desinformação Russa: Em Defesa do Acirramento da Luta Online**" (julho de 2018).⁶ Voltamos ao tópico por causa dos efeitos perniciosos que a desinformação pode ter sobre as eleições. Em muitos casos, a desinformação foi concebida para erodir valores democráticos, aumentar o cinismo e exacerbar a polarização política.

Nosso foco continua sendo as ações que as maiores e mais influentes empresas de mídia social deveriam fazer em relação à desinformação. Isso ocorre porque a alternativa mais óbvia —regulamentação governamental de conteúdo on-line— imediatamente levantaria preocupações sobre liberdade de expressão envolvendo os limites da intervenção estatal e censura. "Os governos só possuem marretas em seus kits de ferramentas", disse Joan Donovan, diretora do Projeto de Pesquisa em Tecnologia e Mudança Social da Harvard Kennedy School, numa entrevista.

A **Primeira Emenda da Constituição dos EUA**, no entanto, não impede que as empresas, o que significa que as empresas de mídia social podem —e costumam fazer— moderar o conteúdo em suas plataformas. Na ausência de regulamentação estatal, cabe às empresas exercer uma governança interna mais rigorosa. Isto é, adotar uma postura mais dura contra a desinformação. Fazer isso representa uma defesa do interesse das próprias empresas. Os governos na Alemanha, França, Austrália e outros países, países que não possuem legislação equivalente à Primeira Emenda, estão aprovando leis para controlar o "discurso de ódio" e as "notícias falsas". Para reduzir a probabilidade de regulamentação de conteúdo feita pelo governo nos EUA, as empresas de mídia social deveriam demonstrar que podem preencher essa brecha de governança, quando se trata de desinformação.

2. A Que Devemos Ficar Atentos em 2020

“
Se os militares dos EUA
pudessem desativar a
Agência Russa de Pesquisa
da Internet pelo menos por
um dia ou dois, isso
significa que as eleições de
2020 serão seguras? Numa
única palavra, não
”

Em 2016, a Agência Russa de Pesquisa da Internet (IRA) se envolveu no que o ex-Conselheiro Especial **Robert S. Mueller III** chamou de "esforços sistêmicos múltiplos para interferir" nas eleições presidenciais dos EUA. ⁷ A IRA permaneceu ativa desde então, mas quando chegaram as eleições minoritárias de novembro de 2018, o volume de intromissão russa era insignificante.⁸ É possível que a IRA tenha tentado acalmar a situação e preparar o terreno para a corrida presidencial de 2020. Também é possível que a remoção mais agressiva de contas suspeitas pelas empresas de mídia social tenha impedido os russos.

E também pode ter havido outro fator: O Comando Cibernético dos EUA, um órgão vinculado ao Pentágono, supostamente usou seus recursos ofensivos de hackers para impedir temporariamente que a IRA utilizasse a Internet no início de novembro de 2018.⁹

Isso significa que as eleições de 2020 serão seguras? Numa única palavra, não. "A IRA é um pequeno componente da operação russa como um todo", que também inclui o serviço de inteligência militar de Moscou e talvez outros, disse **Lee Foster**, que lidera a equipe de desinformação da FireEye, uma empresa de segurança cibernética cujos clientes incluem algumas das principais empresas de mídia social. "Todos esses atores reformulam suas abordagens e táticas", acrescentou.¹⁰ Como resultado, disse **Kevin Carroll**, ex-conselheiro sênior do Departamento de Segurança Nacional, a relativa calma durante as eleições de 2018 não garante que acontecerá o mesmo em 2020.¹¹

O presidente Trump evita reconhecer esse risco, pois ele entende, de acordo com alguns analistas, que dar atenção à interferência da Rússia nas eleições é o mesmo que questionar a legitimidade da sua vitória em 2016.¹² Mas a comunidade de inteligência dos EUA não é tão reticente

assim. Em sua Avaliação Mundial de Ameaças de 2019, a Diretoria de Inteligência Nacional previu que, no próximo ano, a Rússia e outros adversários americanos "certamente usarão operações de influência on-line para tentar enfraquecer instituições democráticas, minar alianças e parcerias dos EUA e definir resultados de políticas".¹³

Deepfakes

Especificamente, a Avaliação Mundial de Ameaças expressou preocupação com vídeos adulterados. Em 2020, afirmou, os adversários dos EUA "provavelmente tentarão usar deepfakes ou tecnologias semelhantes de aprendizagem de máquina para criar arquivos de imagem, áudio e vídeo convincentes —mas falsos— para aumentar as campanhas de influência contra os Estados Unidos e nossos aliados e parceiros".

O termo deepfakes vem de uma combinação de "aprendizagem profunda" e "falsificações". Teve origem em 2017 com um usuário anônimo do Reddit que se chamava de "deepfakes". Esse indivíduo ganhou atenção usando algoritmos de aprendizagem profunda para sobrepor rostos de celebridades nos corpos de atores pornográficos. Sob pressão externa, o Reddit proibiu os deepfakes individuais e pornográficos em geral, mas os imitadores proliferaram em outros lugares da Internet.¹⁴

“
Deepfakes podem ser produzidos por qualquer pessoa com um computador, acesso à Internet e interesse em influenciar uma eleição.’—
John Villasenor, Professor de Engenharia Elétrica, Políticas Públicas e Gerenciamento da University of California, Los Angeles

A indústria cinematográfica alterou imagens de vídeo por décadas. Mas os truques de Hollywood exigem técnicos qualificados, muito dinheiro e muito tempo. Ferramentas de IA com código aberto democratizaram a fabricação de vídeos. "Deepfakes podem ser feitos por qualquer pessoa com computador, acesso à Internet e interesse em influenciar uma eleição", segundo **John Villasenor**, Professor de Engenharia Elétrica, Políticas Públicas e Gerenciamento da University of California, Los Angeles.¹⁵

Uma breve digressão para algumas informações técnicas: O termo **inteligência artificial** (IA), de modo geral, se refere a algoritmos capazes de executar tarefas humanas, como o reconhecimento de um rosto humano. A aprendizagem profunda, um subconjunto da IA, se refere a certos tipos de algoritmos que podem aprender e tomar decisões inteligentes por conta própria. Deepfakes são vídeos falsificados, feitos por meio de aprendizagem profunda.

Um sistema de aprendizagem profunda pode produzir uma falsificação persuasiva estudando fotografias e vídeos de uma pessoa-alvo, bem como o vídeo de um ator falando e se comportando da maneira que o alvo será representado no novo vídeo falso. Em suma, a tecnologia é capaz de fundir o alvo e o ator. Depois que a versão preliminar de um vídeo falso é produzida, um método conhecido como GANs, ou redes adversárias generativas, é responsável por torná-lo mais crível. GANs procura detectar falhas no vídeo adulterado, melhorando o tratamento de falhas. Após várias rodadas de detecção e aprimoramento, o deepfake fica pronto.

Em 7 de junho de 2019, dois artistas britânicos de vídeo postaram no Instagram um **deepfake do Mark Zuckerberg**, fundador e CEO do Facebook. O vídeo foi preparado por uma startup de tecnologia israelense chamada Canny AI. A empresa usou como alvo um vídeo que o Mark Zuckerberg divulgou em 2017. Canny AI disse, por e-mail, que alterou os movimentos faciais do vídeo original para coincidir com os de um ator cuja voz foi substituída pela de Zuckerberg. O resultado

é visualmente convincente, mas a voz do ator não é parecida com a do executivo de tecnologia. Além disso, os artistas sinalizaram sua intenção de satirizar Zuckerberg, fazendo-o falar como um super-vilão de James Bond: "Imagine isso por um segundo: um homem com controle total de dados roubados de bilhões de pessoas, todos os seus segredos, vidas, futuro".¹⁶

É surpreendente que deepfakes mais sérios ainda não tenham surgido na política americana. Especialistas ouvidos numa audiência do Comitê de Inteligência da Câmara em junho de 2019 afirmaram que é apenas uma questão de tempo. "Imagine se na noite anterior às eleições de 2020, numa eleição apertada, um deepfake mostrasse um candidato fazendo algo chocante que ele nunca fez", sugeriu **Danielle Keats Citron**, professora de direito da Boston University. "O dano seria irreparável. O resultado das eleições não pode ser desfeito."¹⁷ (Uma suposta deepfake do presidente do Gabão, **Ali Bongo**, fazendo um discurso de ano novo em 2019 ajudou a precipitar um golpe de estado malsucedido no país africano.¹⁸

Citron e dois coautores, escrevendo para o site Lawfare, observaram que a difamação política nos EUA remonta à rivalidade entre Alexander Hamilton e Thomas Jefferson. "Diferente hoje", disseram eles, "é que as falsidades envolvem 'provas' visuais e sonoras em que nossos olhos e ouvidos estão profundamente inclinados a confiar (não apenas palavras escritas que podem ser descartadas com mais facilidade)." E graças às redes sociais, "as fraudes podem chegar rapidamente a inúmeros indivíduos".¹⁹

Alguns estudiosos da desinformação subestimam o potencial dos deepfakes, argumentando que fraudes menos elaboradas —como os vídeos da Nancy Pelosi ou as imagens grosseiramente adulteradas em Photoshop— podem ser suficientes para desestabilizar uma eleição. "Não precisamos de deepfakes. Muitas pessoas não precisam de provas 'reais'", afirmou Patrick Warren, um economista da Clemson University que estuda redes sociais. Em uma nação politicamente

polarizada, acrescentou, "trata-se de ideologia e de encontrar coisas que confirmam aquilo em que [certos grupos de eleitores] já acreditam". 20

Warren pode ter razão quando se trata de eleitores mais engajados. Mas sua observação não nega efeito potencial de uma série de vídeos incriminadores e aparentemente genuínos divulgados na véspera do dia da votação.

Americanos Desavisados

Diferentemente dos produtos de última geração, baseados em IA, alguns métodos de desinformação têm por objetivo a tradicional mobilização para levar as pessoas às ruas. "Um motivo de maior preocupação é o esforço para recrutar americanos de verdade para, de modo desavisado, organizar comícios políticos ou criar novos grupos on-line", disse **Laura Rosenberger**, diretora da Aliança para Garantia da Democracia, numa entrevista.

O **relatório final** de março de 2019, do Conselheiro Robert Mueller, jogou luz sobre a atuação dos agentes da IRA que, passando-se por ativistas genuínos dos EUA, foram capazes de mobilizar cidadãos americanos a participar de dezenas de comícios a partir de 2016. Ao expandir essa estratégia em 2020, os russos atingiriam um dos seus principais objetivos —converter influência no mundo digital em discórdia no mundo real.

Em 2016, a técnica envolveu perfis ou grupos falsos em redes sociais da IRA, anunciando eventos contrários e a favor dos muçulmanos, que os russos organizaram com sucesso em maio de 2016, em Houston. A IRA enviava mensagens diretas de divulgação dos eventos para os seus seguidores americanos online. Daqueles que indicaram interesse em participar, a IRA passou a recrutar pessoas locais para coordenar os eventos. Os agentes da IRA dariam uma desculpa para não comparecer, mas buscavam a cobertura da mídia convencional. O número de participantes nos eventos variava, de alguns poucos, até centenas de pessoas. 21

Um incidente no verão de 2018 serve para ilustrar o resultado dessa estratégia. Uma conta falsa do Facebook, de inclinação à esquerda, chamada de Resisters, que o **Facebook identificou** como sendo possivelmente da Rússia, criou um evento em 10 de agosto para contrapor um encontro organizado por grupo supremacistas brancos em Washington, DC. Resisters, a primeira página do Facebook a promover o contraprotesto, passou a atuar em conjunto com outras cinco páginas aparentemente reais dedicadas à oposição aos supremacistas brancos, de acordo com o Facebook. "Essas páginas legítimas, de modo inadvertido, ajudaram a aumentar o interesse na [demonstração contrária] e publicaram informações sobre transporte, materiais e locais para que as pessoas pudessem participar dos protestos", acrescentou a empresa. Antes das manifestações, no entanto, o Facebook determinou que a resistência era uma fraude e a removeu. A empresa também notificou 2.600 usuários que indicaram ao Resisters o interesse em participar. 22

No final, apenas 40 supremacistas brancos apareceram para a manifestação perto da Casa Branca, e foram superados e muito pelo número de manifestantes contrários. 23 Apesar de ter sido descoberto, os Resisters alcançaram seu objetivo de amplificar a divisão na sociedade americana.

Instagram

O Instagram não recebeu tanta atenção no contexto da desinformação quanto o Facebook, Twitter e YouTube, mas teve um papel muito maior na manipulação eleitoral da Rússia em 2016 do que a maioria das pessoas imagina. E pode se tornar um instrumento crucial para a Rússia no próximo ano, de acordo com um relatório encomendado pelo Comitê de Inteligência do Senado. 24

Iniciado em 2010, o Instagram foi adquirido pelo Facebook depois de 18 meses, por US\$ 1 bilhão. Hoje, o Instagram tem 1 bilhão de usuários (comparado com 2,38 bilhões do Facebook, 2 bilhões do YouTube e 330 milhões do Twitter). A plataforma de postagem de fotos e vídeos é um destino comum para usuários mais jovens entediados com o Facebook. 25

“

O fato de o Instagram ter superado o Facebook como uma plataforma de engajamento utilizada pela Rússia pode 'indicar sua força como uma ferramenta na guerra de memes centrada na imagem', de acordo com um relatório encomendado pelo Comitê de Inteligência do Senado.

”

O serviço orientado a imagem do Instagram o torna o local ideal para memes, que são fotos combinadas com um texto curto e contundente. Os memes, por sua vez, são um veículo popular para citações falsas e outras desinformações.

Em 2016, a IRA alcançou mais envolvimento do usuário no Instagram do que em qualquer outra rede social. Essa foi uma descoberta do **relatório** do Comitê de Inteligência divulgado em dezembro de 2018. "O Instagram foi uma frente significativa na operação de influência da IRA, algo que os executivos do Facebook parecem ter evitado mencionar no depoimento ao Congresso", afirmou o relatório. "Pela nossa avaliação, o Instagram provavelmente será o principal campo de batalha nos próximos tempos." O relatório foi escrito para o comitê por especialistas do Tow Center de Jornalismo Digital da Columbia University e por duas empresas de pesquisa, New Knowledge e Canfield Research.

Analisando dados de 2015 a 2018, os pesquisadores descobriram que havia 187 milhões de envolvimento de usuários com material da IRA no Instagram - mais do que o dobro do Facebook (77 milhões) ou Twitter (73 milhões). Um meme postado no Instagram pela conta "Blacktivist" da IRA, mostrou um policial parcialmente vestido com um capuz e traje da Ku Klux Klan acima da frase: "Faz anos que a KKK está infiltrada nos departamentos de polícia". 26

Muitos americanos dizem que a desinformação é um problema crítico que precisa ser corrigido.

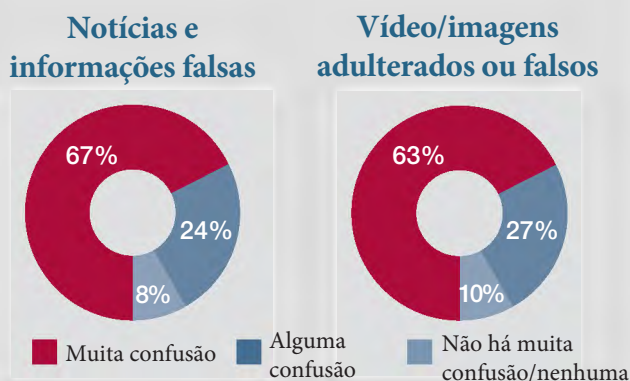
Percentual de americanos que afirma que esse é um problema muito grande para o país na atualidade

70%	Toxicodependência
67%	Acessibilidade da assistência médica
52%	Sistema político dos EUA
51%	Diferença entre ricos e pobres
50%	Notícias/Informações Falsas
49%	Crime violento
46%	Mudança climática
40%	Racismo
38%	Imigração ilegal
34%	Terrorismo
26%	Sexismo

Porcentual que diz que as notícias e as informações inventadas têm um grande impacto sobre...

Confiança dos americanos no governo	68%
Confiança dos americanos	54%
Capacidade dos líderes políticos de realizar o trabalho	51%

Dois terços dos adultos americanos acham que as informações falsas e os vídeos adulterados criam muita confusão sobre questões e eventos da atualidade.



Observação: Foram desconsiderados os que não responderam. Fonte: Pesquisa do Centro de Pesquisa Pew envolvendo 6.127 adultos nos EUA, entre 19 de fevereiro e 4 de março de 2019

Porcentual de americanos que tem pelo menos "alguma" confiança nas redes sociais

42%	YouTube	
29%	Facebook	
23%	Instagram	
22%	Twitter	
16%	Snapchat	

Porcentual que acredita que as redes sociais são pelo menos "um pouco" responsáveis por espalhar a desinformação

64%	Facebook	
55%	Twitter	
48%	YouTube	
46%	Instagram	
39%	Snapchat	

Source: Institute for Public Relations survey of 2,200 Americans between March 19 and March 24, 2019

O fato de o Instagram ter superado o Facebook como uma máquina de envolvimento da IRA pode "indicar sua força como ferramenta na guerra de memes centrada na imagem", concluiu o relatório do comitê. Também é possível, reconheceram os pesquisadores, que "fazendas de clique" foram responsáveis pela criação de envolvimento no Instagram da IRA. Mas isso não alterou a previsão dos pesquisadores de que o Instagram teria um papel importante em 2020.

Outros observadores concluíram que, além da interferência russa, hoaxes e teorias de conspiração geradas no próprio país crescem cada vez mais no Instagram. Um exemplo é o QAnon, uma teoria da conspiração de direita amplamente divulgada sobre um plano de "estado profundo", para derrubar o presidente Trump. Em março de 2019, *The Atlantic* previu que o Instagram "provavelmente é onde será travada a próxima grande batalha contra a desinformação, mas essa plataforma tem em grande parte conseguido escapar de qualquer controle". 27 A aproximação das próximas eleições vai exigir um exame minucioso.

WhatsApp

Pesquisadores do **Laboratório de Pesquisa sobre Crimes Digitais**, vinculado ao Atlantic Council, sugeriram numa entrevista que um outro veículo do Facebook, o WhatsApp, merece destaque em nossa lista de atenção. O motivo da preocupação com o WhatsApp se baseia em duas observações: Primeiramente, o serviço de mensagens é criptografado "de ponta a ponta", o que significa que mesmo os supervisores da empresa não conseguem ver o que os usuários estão fazendo. Isso é ótimo para privacidade, mas também é uma receita para fazer irregulares ou para coisas pior.

Em segundo lugar, o WhatsApp serviu como um importante veículo para disseminar informações falsas durante as recentes eleições presidenciais no Brasil e na Índia. As pessoas nesses países usavam o recurso de encaminhamento do WhatsApp para espalhar conteúdo extremamente enganador —grande parte dele gerado por campanhas em guerra— para grandes grupos de usuários. O conteúdo do WhatsApp pode ter mais credibilidade aos

olhos dos destinatários, porque geralmente vem de remetentes com quem os destinatários têm familiaridade. No Brasil, o candidato presidencial de extrema direita, Jair Bolsonaro, sagrou-se vencedor em outubro de 2018, "em parte, com a ajuda de uma onda de desinformação e de rumores tóxicos, grande parte espalhada pelo WhatsApp", informou o Financial Times. Sete meses depois, Narendra Modi ganhou a reeleição como primeiro-ministro da Índia em uma eleição marcada por falsidades, em que o WhatsApp se tornou "o campo central de batalha". 28

No entanto, há razões para duvidar que o WhatsApp será amplamente utilizado para causar problemas nos EUA. Adquirida pelo Facebook em 2014 por US\$ 22 bilhões, a plataforma de mensagens ganhou enorme popularidade em muitas partes do mundo. Possui 1,5 bilhão de usuários em todo o mundo, 300 milhões deles na Índia e outros 120 milhões no Brasil. Mas não alcançou a mesma penetração nos EUA, onde possui menos de 60 milhões de usuários adultos. Sua menor base de usuários nos EUA limita seu potencial para disseminar desinformação. Além disso, os americanos que usam o WhatsApp, o fazem principalmente para comunicação de mensagens individuais, ou mensagens de uma pessoa para poucas outras, e não para muitas, como ocorreu durante as eleições no Brasil e na Índia. Com as mensagens individuais, a desinformação (ou qualquer tipo de conteúdo) tem muito menos probabilidade de se tornar viral. Por fim, o WhatsApp limitou o número de grupos aos quais um usuário pode encaminhar uma parte do conteúdo de uma só vez —uma ação pensada para retardar a distribuição de inverdades.

Ainda assim, o Laboratório de Pesquisa sobre Crimes Digitais adverte que o WhatsApp não deve ser ignorado. "Já se sabe ao certo que foi utilizado [para desinformação] em outros países", disse Iain Robertson, editor-chefe substituto do laboratório. "Existem margem para que isso se expanda aqui." 29

Irã e China

Inevitavelmente, outros rivais estrangeiros dos EUA se basearam na operação de desinformação promovida pela Rússia para lançar a sua. O que tem a maior chance de aparecer durante as eleições de 2020 é o Irã, especialmente se a hostilidade entre Teerã e Washington permanecer aguçada.

“

‘ Neste momento, os iranianos estão seguindo a cartilha de operações do Kremlin', disse o senador Mark Warner (D., Va.) em agosto de 2018, na época em que o Facebook eliminou 652 contas, páginas e grupos falsos ligados com origem no Irã e na Rússia.

”

O Irã tem sido um dos inimigos digitais mais consistentes dos EUA na última década, mais conhecido por ataques cibernéticos a bancos, hospitais e agências governamentais. Nos últimos anos, também testou seu aparato de desinformação direcionado aos EUA. Em maio de 2019, a FireEye, empresa de segurança cibernética, divulgou um **relatório** identificando um conjunto de contas em redes sociais baseado no Irã que usava perfis americanos falsos para defender visões políticas conservadoras e liberais. Algumas das contas falsas simulavam candidatos políticos republicanos verdadeiros, que participaram das eleições de 2018. A rede iraniana promoveu temas anti-sauditas, anti-israelenses e pró-palestinos. Também expressou apoio ao acordo nuclear de 2015, do qual o presidente Trump se retirou unilateralmente. Alguns dos perfis falsos utilizaram nomes como "John Turner" e "Ed Sullivan" —tiveram cartas publicadas sobre esses temas no New York Daily News, Los Angeles Times e em outros meios de comunicação tradicionais. 30

Em resposta ao FireEye, o Facebook removeu quase 100 contas, páginas e grupos do Facebook e Instagram. 31 Agindo de forma independente, o Twitter anunciou que havia removido mais de 2.800 "contas não autênticas com origem no Irã". 32 Essas ações seguiram outras ações do Facebook para eliminar contas iranianas em 2018. "Os iranianos estão agora seguindo a

“
 Umadasteoriasde
 conspiraçãodeorigemlocal
 queprovavelmente
 ganharãoforçanos
 próximosmeseséqueas
 principaisempresasde
 mídiasocialestão
 conspirandocomos
 democratasparaderrotara
 candidaturadopresidente
 Trumpàreeleição.”

Cartilha do Kremlin", disse o senador **Mark Warner** (D., Va.) em agosto de 2018, na época em que o Facebook eliminou 652 contas, páginas e grupos falsos ligados com origem no Irã e na Rússia. 33

O Irã e a Rússia não são os únicos adversários dos EUA ou possíveis concorrentes estratégicos que serão fonte de desinformação em 2020. Por estar envolvida numa dura guerra comercial com os EUA, a China também pode representar uma ameaça digital.

O fato de haver tarifas mais altas dos EUA ou outras penalidades contra a economia chinesa pode ser motivo suficiente para Pequim desencadear "um exército de agentes" para atuar no ambiente online em 2020, de acordo com **Dipayan Ghosh**, ex-consultor de políticas do Facebook e diretor do Projeto Platform Accountability na Harvard Kennedy School. 34 Kevin Carroll, ex-conselheiro sênior de Segurança Nacional, concordou, dizendo: "Se as relações entre a China e os Estados Unidos se deteriorarem ainda mais, os chineses podem atacar as redes sociais americanas utilizando simplesmente a força bruta, com ou sem o incentivo dos serviços de segurança chineses, já que muitos chineses bilíngues e nacionalistas ávidos estão ativos online." 35

Uma grande diferença entre as possíveis campanhas digitais da Rússia e da China: Os russos apoiaram Trump, observou Ghosh, mas "os chineses provavelmente trabalhariam para tirá-lo do cargo, porque não gostam das suas políticas ou da sua retórica".

Desinformação Doméstica

Embora a interferência eleitoral estrangeira tenha dominado a discussão atual sobre desinformação, o conteúdo falso mais malicioso nos EUA é gerado dentro do próprio país. Existe esse consenso no meio dos executivos das empresas de mídia social, detetives de segurança cibernética e pesquisadores acadêmicos. 36 Referindo-se à variedade local mais comum, **Justin Hendrix**, diretor executivo do NYC Media Lab, disse numa entrevista: "Acho que veremos muitíssima coisa desse tipo" em 2020.

Matt Masterson, consultor sênior de segurança cibernética do Departamento de Segurança Nacional, observou recentemente durante uma conferência organizada pela Fundação Hewlett, que na opinião do governo federal a desinformação gerada no país é "o maior desafio que enfrentamos". 37 A interferência estrangeira numa eleição nos EUA é claramente ilegal. Os EUA indiciaram cerca de duas dúzias de agentes da IRA e do serviço de inteligência militar da Rússia, embora seja altamente improvável que Moscou extradite qualquer um dos réus. Por outro lado, é difícil afirmar que o conteúdo criado por cidadãos dos EUA é enganoso, e não expressão política geral, que goza de proteção da Primeira Emenda contra regulação ou controle por parte do governo.

A desinformação local produzida tanto pela esquerda quanto pela direita, mas, como indicado em nosso relatório de março de 2019, os estudos mostram que os usuários conservadores das redes sociais têm maior probabilidade de distribuir conteúdo falso. 38 Uma das teorias de conspiração geradas localmente que provavelmente ganharão força no meio dos grupos de direita nos próximos meses é que as principais empresas de mídia social estão conspirando com os democratas para derrotar o presidente Trump em sua campanha à reeleição. Breitbart News, um importante canal de extrema direita com 3,87 milhões de seguidores no Facebook, divulgou a alegação de que "grandes empresas de tecnologia estão tentando garantir que vozes conservadoras sejam suspensas antes do ciclo eleitoral de 2020". 39 No ano passado, Facebook, Instagram, Twitter e YouTube proibiram uma série de personalidades da direita, incluindo Alex Jones, proprietário do site de conspiração Infowars. Para justificar essas exclusões, as redes sociais alegam que ocorreu o descumprimento das suas políticas contra a promoção do ódio ou da violência, e negam qualquer vingança contra os conservadores. 40

Mas o presidente Trump não aceita essa alegação. Em uma série de tweets e entrevistas, ele atacou as grandes empresas de tecnologia. "Posso dizer que eles me discriminam", disse à CNBC. "Você sabe, as pessoas falam sobre conluio. O verdadeiro conluio é entre os democratas e essas empresas." 41

Apresentando uma "cúpula de rede social" de direita na Casa Branca em julho de 2019, Trump twittou sobre "a tremenda desonestidade, preconceito, discriminação e supressão praticados por certas empresas". Em tom de ameaça, disse: "Não vamos deixar que eles se safem por muito mais tempo." 42 O presidente disse aos participantes que estava orientando seus assessores a explorar "todas as soluções regulatórias e legislativas" contra a suposta trama do Facebook, Google e Twitter. 43

O propósito aparente dessa reunião na Casa Branca foi o de incentivar fornecedores de desinformação conservadores que, mesmo antes de Trump lhes dar seu selo público de aprovação, começaram a perseguir possíveis candidatos democratas à Presidência. Alguns levantaram questões espúrias sobre a senadora **Kamala Harris**, da Califórnia, se ela realmente é afrodescendente. 44 Outros orquestrado acusações de agressão sexual falsas contra South Bend, Ind., contra o prefeito **Pete Buttigieg**. 45

Também houve desinformação produzida localmente pela esquerda. Uma entrevista de 11 de janeiro de 2019 na estação de rádio pública WNYC tratou de um esquema liberal de distorção de notícias pelas redes sociais no Alabama. "Essa foi uma operação de bandeira falsa", disse o ativista democrata **Matt Osborne**. "Essa ideia teve origem na documentação sobre o que foi feito pelos russos". 46

Osborne ajudou a criar uma página falsa no Facebook e uma conta no Twitter chamada de "Dry Alabama", que parecia ter sido criada por eleitores conservadores batistas que apoiavam o republicano Roy Moore numa eleição especial para o Senado dos Estados Unidos.

O objetivo era associar Moore a uma medida de proibição ao álcool no Estado e, assim, alienar republicanos moderados e defensores da livre iniciativa. O esquema foi realizado em nome do democrata Doug Jones, que acabou vencendo a eleição, mas, segundo os relatos, não tinha conhecimento dessa atividade digital. 47 Supostamente, os democratas executaram outra manobra de bandeira falsa, fazendo uso de mais de uma dúzia de páginas enganosas do Facebook, destinadas a atrair conservadores em todo o país na véspera das eleições minoritárias de 2018. 48

Alegando que os republicanos estão usando táticas enganosas semelhantes, Osborne disse que os democratas devem combater fogo com fogo.

Em 2020, ele previu, "você verá um movimento de gastos espúrios em campanhas digitais nos últimos dias da corrida". 49 Por exemplo, ele incluiu numa troca de mensagens por email: "A dissuasão pode ser apenas para mostrar, por exemplo, a um eleitor republicano a imagem de uma onda vermelha, com uma mensagem de vitória garantida: 'Você nem precisa votar. Trump já ganhou.'" 50

Empresas com Fins Lucrativos

Outra pessoa que participou do episódio de desinformação do Alabama foi **Jonathon Morgan**, CEO da New Knowledge, uma empresa de pesquisa de rede social com sede em Austin, Texas. Seu trabalho aponta para um perigo adicional em 2020: o envolvimento de empresas privadas ou de seus executivos na elaboração de desinformação.

Uma vez que a grande mídia trouxe à tona o caso do Alabama, Morgan disse que estava atuando como pesquisador, buscando entender como funciona a desinformação digital, para não afetar o resultado da eleição ao Senado envolvendo os candidatos Jones e Moore. No entanto, o Facebook suspendeu a conta de Morgan e outras quatro administradas por pessoas envolvidas no episódio. A atividade no Alabama foi financiada por doadores políticos democratas, incluindo Reid Hoffman, co-fundador do LinkedIn, que se desculpou publicamente e disse que não sabia das táticas maliciosas. 51

Convidado a se manifestar, Morgan disse que "não era nossa intenção criar uma página enganosa". Sua empresa, New Knowledge, foi mencionada anteriormente neste relatório porque ajudou a elaborar uma avaliação sobre a interferência russa para o Comitê de Inteligência do Senado.

Na Rússia, os serviços de desinformação são vendidos sem qualquer controle ou motivação. Jigsaw, um think tank corporativo afiliado ao Google, decidiu testar o mercado russo ao contratar uma empresa chamada de **SEO Tweet**,

“

'Esta foi uma operação de bandeira falsa. Essa ideia teve origem na documentação sobre o que foi feito pelos russos.' — ativista democrata Matt Osborne, descrevendo um complô liberal para distorção das redes sociais no Alabama

”

descoberta a partir de campanhas publicitárias online. O preço de uma modesta campanha de desinformação de duas semanas destinada a um site fictício que o Jigsaw criou especialmente para esse fim foi de meros US\$ 250. A transação, realizada na primavera de 2018, foi um experimento projetado para ver como as operações de influência poderiam ser compradas na Rússia, disse Jigsaw à Wired. 52 SEO Tweet não respondeu a um pedido de comentário enviado via Twitter.

Israel, que abriga um mercado de tecnologia robusto, também poderia ser fonte de campanhas de desinformação com fins lucrativos. O deepfake satírico do Mark Zuckerberg foi criado por uma start-up israelense. Num episódio independente, um artigo de fevereiro de 2019 no **The New Yorker** descreveu uma empresa de inteligência israelense chamada Psy-Group que se especializou em "atuar secretamente para espalhar mensagens e influenciar as convicções e o comportamento das pessoas". Entre as ações atribuídas à empresa, estavam duas campanhas digitais nos EUA: uma teve como alvo estudantes universitários americanos que defendiam o boicote a Israel, e uma segunda que tentou virar uma eleição para o conselho hospitalar da Califórnia. Em 2016, o Psy-Group ofereceu sem sucesso os seus serviços para a campanha de Trump. A empresa foi fechada em 2018 e as pessoas até então associadas a ela não responderam aos convite para se manifestarem. 53

O Facebook anunciou em maio de 2019 que estava agindo contra outra instituição israelense, uma empresa de consultoria e lobby chamada **Archimedes Group**, cujos funcionários supostamente realizaram campanhas de desinformação política na África e em todo o mundo. Os agentes se disfarçaram de pessoas locais, incluindo organizações de notícias locais, e publicaram informações sobre supostos vazamentos a respeito de políticos africanos. Esse foi o relato do Facebook depois de eliminar 265 contas, páginas, grupos e eventos do Archimedes. Supostamente, a atividade enganosa se concentrou na Angola, Níger, Nigéria, Senegal, Togo e Tunísia e também se estendeu à América Latina e ao Sudeste Asiático. Archimedes teria gasto US\$ 812.000 em publicidade no Facebook, pagos em dólares americanos, shekels israelenses e reais brasileiros. Aparentemente, o Facebook não descobriu se o Archimedes tinha uma agenda mais ampla, e não há provas de que a atividade estivesse ligada ao governo de Israel. "Não queremos que nossos serviços sejam usados para manipular pessoas", escreveu o chefe de segurança cibernética do Facebook, Nathaniel Gleicher, num post da empresa. 54

Não foi possível estabelecer contato com a empresa nem com os executivos do Archimedes. Antes do desaparecimento do seu site devido à ação do Facebook, a empresa se descreveu como tendo "todas as vantagens disponíveis para mudar a realidade de acordo com os desejos dos nossos clientes". 55

Supressão do Eleitor Digital

Uma categoria final de desinformação que quase certamente surgirá em 2020 é a supressão de eleitores. Não há nada de novo em convencer certos eleitores a não irem às urnas, mas a difusão da rede social torna essa prática particularmente perniciosa quando realizada nas plataformas digitais.

Existem diversos exemplos recentes de tentativas de supressão de eleitores nas redes sociais. Como parte de sua campanha a favor do Trump em 2016, os agentes russos da IRA tentaram convencer os afrodescendentes dos Estados Unidos de que não havia sentido em votar porque os

democratas não se importavam com eles. 56 Facebook disse, algumas semanas antes das eleições minoritárias de 2018, que encontrou e eliminou 45.000 peças de conteúdo voltadas à supressão de eleitores. 57

Pesquisadores da **University of Wisconsin** encontraram várias categorias de supressão pela rede social no segundo semestre. Algumas postagens no Twitter tentaram enganar os eleitores contrários ao presidente Trump, fornecendo informações incorretas sobre o dia da votação ou anunciando que "2018 é o primeiro ano em que os cidadãos podem votar por MENSAGEM DE TEXTO." Outros tweets tentaram intimidar os eleitores liberais dizendo aos membros da Associação Nacional de Armas e aos republicanos que levassem suas armas para as urnas. No entanto, outros posts de rede social tentaram afastar os eleitores negros em 2018, alegando que a votação seria inútil ou que não haveria diferença entre os dois principais partidos. Muitas dessas mensagens usavam hashtags como #dontvote e #dontbeavoter [#não_vote, #não_seja_um_eleitor]. 58

Algumas ações voltadas à supressão de eleitores utilizaram argumentos mais específicas. O Twitter disse que desativou milhares de robôs anônimos de supressão de eleitores em 2018, incluindo um grupo com mais de 10.000 robôs disfarçados para parecer que eram democratas. Alguns dos tweets tentavam convencer os democratas do sexo masculino a não participar das eleições minoritárias, do contrário, os homens abafariam as vozes das mulheres. 59

Poucas medidas atingem tão de cheio o cerne da democracia como as voltadas à confusão ou intimidação de pessoas com direito a voto. As empresas de mídia social terão que permanecer atentas a esse respeito.

“

Poucas medidas atingem tão de cheio o cerne da democracia como as voltadas à confusão ou intimidação de pessoas com direito a voto. As empresas de mídia social terão que permanecer atentas a esse respeito.

”

3. Qual a Resposta Dada por Empresas de Mídia Social a Ações de Desinformação



As empresas estão melhorando muito na detecção e remoção de contas falsas." - Dipayan Ghosh, diretor do Projeto Platform Accountability da Harvard Kennedy School



De certa forma, as principais empresas de mídia social estão se preparando desde 2017 para as eleições de 2020, quando começaram a reconhecer as ações de desinformação promovidas pela Rússia. Desde então, essas empresas adotaram diversas medidas para limpar os seus sites e fortalecer suas defesas. Eles também implementaram uma série de medidas especificamente destinadas a proteger a integridade das eleições. Ambas as categorias — medidas gerais e medidas específicas para as eleições — devem influenciar o cenário de 2020.

É preciso reconhecer que as empresas estão se comunicando mais entre si, com o governo e com especialistas externos; estão eliminando mais contas fraudulentas; está havendo uma maior mobilização de equipes especialmente dedicadas a combater irregularidades nas eleições. Mas ainda há muito a fazer.

Mudanças Gerais

Eliminação de Contas Falsas

"As empresas estão melhorando muito na detecção e remoção de contas falsas", disse Dipayan Ghosh, co-diretor do Projeto Platform Accountability de Harvard, em entrevista. O Facebook emprega ferramentas aprimoradas de IA para excluir contas falsas de forma automática. O número está na casa dos bilhões — 2,19 bilhões nos primeiros três meses de 2019. A maioria dessas contas foi bloqueada poucos minutos depois da sua criação, impedindo-as de causar danos, de acordo com a empresa. 60

Nos últimos dois anos e meio, o Facebook também anunciou dezenas de ações de pequena escala, mas bastante específicas, voltadas à eliminação de milhares de contas e páginas que demonstraram "comportamento coordenado não autêntico". 61 Na opinião do Facebook, essas contas e páginas trabalharam em conjunto para enganar os usuários sobre a

identidade dos seus responsáveis e sobre o que estão fazendo. A empresa enfatiza que está buscando coibir comportamentos enganosos, não suprimir conteúdo. Mas o comportamento em questão geralmente inclui ações voltadas à disseminação de desinformação. O Twitter, por sua vez, eliminou milhões de contas falsas por razões semelhantes. YouTube, também, ainda que em menor grau.

Em comparação à relativa passividade das empresas em 2016, essas ações parecem ser mais rigorosas. Além de aperfeiçoar seus algoritmos de triagem de IA, as plataformas contrataram milhares de revisores adicionais e prestadores de serviço para vasculhar contas que exibiam conteúdo problemático. Mas o número impressionante de contas falsas ou enganosas eliminadas também indica a vasta oferta dessas contas e a quase certeza de que as empresas não estão capturando todas elas — talvez nem a maioria delas. "Muitas contas falsas não são detectadas e podem ser usadas para manipulação", de acordo com Filippo Menczer, professor de informática e ciência da computação na Indiana University, que estuda desinformação. Essas contas operadas por robôs, acrescentou ele numa troca de mensagens de e-mail, "podem ser usadas para ampliar a disseminação de desinformações, deepfakes, ataques, confusões, supressão de eleitores ou falso apoio a candidatos". 62

Como Identificar e Eliminar Conteúdo Falso

Junto com o problema de contas falsas e de comportamento não autêntico está a existência generalizada de conteúdo falso que se espalha rapidamente nas redes sociais. O que deve acontecer com as notícias falsas quando contas e páginas legítimas começarem a compartilhá-las? Para lidar com as falsidades que podem distorcer o discurso político do país, as plataformas investiram em ferramentas mais precisas de IA, expandiram a análise interna e adicionaram mais recursos independentes de verificação de fatos. O Facebook agora envia histórias e imagens potencialmente falsas para 53 organizações de verificação de fatos envolvendo 42 idiomas em todo o mundo. Em particular, a rede social enfatizou que está "usando tanto tecnologia como pessoas para combater o aumento da informação falsa baseada em fotos e vídeos" — um compromisso que reflete a crescente preocupação com os deepfakes. 63

Mark Zuckerberg, do Facebook, disse recentemente que sua empresa está considerando banir e eliminar todos os deepfakes, pelo menos os que for capaz de identificar. 64 Mas, geralmente, as plataformas não removem o conteúdo simplesmente porque ele é considerado falso. Em vez disso, elas geralmente reduzem a importância do material falso, com base na noção de que seus executivos e funcionários não devem ser "árbitros da verdade". 65 Facebook, por exemplo, "classifica 80% das notícias falsas nos Feeds de Notícias dos usuários, classifica o conteúdo como falso e fornece contexto sobre o tópico. Foi isso que o Facebook fez com o vídeo adulterado fazendo com que a Nancy Pelosi parecesse estar bêbada. 66 As empresas, porém, não aplicam esses procedimentos de rebaixamento de conteúdo de maneira consistente. O Facebook disse que, em certos contextos, elimina o conteúdo falso. Os exemplos incluem informações falsas que representam risco de violência física ou têm como objetivo afastar os eleitores das urnas, ou que buscam interferir no censo

dos EUA. 67 Twitter e YouTube disseram que também eliminarão o conteúdo voltado à supressão de eleitores. 68

Em junho de 2019, o YouTube anunciou que começaria a excluir alguns conteúdos por considerá-los enganosos. Em outras palavras, em certas circunstâncias, atuaria como árbitro da verdade. Especificamente, o YouTube disse que eliminaria vídeos "negando a existência de eventos violentos bem documentados, como o Holocausto ou tiroteio na escola Sandy Hook Elementary". 69 Vídeos que perpetuam outras falsidades, como a afirmação de as imagens da chegada do homem à Lua com a Apollo foram encenação, ou que uma cura milagrosa falsa ajudará pacientes com uma doença grave, não são mais recomendados aos usuários do YouTube, mas permanecem disponíveis na plataforma. 70 Não existe um princípio óbvio para embasar essas distinções.



Reconhecendo que precisa encontrar outras maneiras para identificar desinformação, o Facebook lançou uma iniciativa de crowdsource para verificação de fatos.



Como foi dito, o NYU Stern Center acredita que, quando as redes sociais descobrem conteúdo comprovadamente falso, especialmente na esfera política, elas não devem apenas reduzir sua visibilidade ou apenas impedir que um algoritmo de recomendação sugira esse conteúdo para os usuários. A plataforma deve remover o material completamente, mantendo apenas uma cópia de referência num repositório de conteúdo falso, já excluído. Esse material deve permanecer acessível ao público, para pesquisadores e usuários em geral, mas não deve circular por feeds ou recomendações. O Twitter deu um passo nessa direção, ao lançar conjuntos de dados substanciais de tweets, que foram removidos após vincular o conteúdo a operações de desinformação apoiadas pela Rússia ou por outros países. 71

Crowdsourcing Verificações de Fatos

Reconhecendo que precisa encontrar maneiras adicionais de identificar desinformação, o Facebook lançou uma iniciativa de crowdsource para verificação de fatos. Com um bilhão de postagens por dia, o volume de conteúdo falso na rede social ultrapassa muito a capacidade de revisores internos e externos. Para lidar com esse desafio, o Facebook diz que está estudando formas para aproveitar "abordagens comunitárias de combate à desinformação". Com isso, a empresa quer dizer "fazer uso de grupos de pessoas que usam o Facebook para apontar fontes jornalísticas que possam corroborar ou contradizer as alegações feitas em conteúdo potencialmente falso". 72

Mark Zuckerberg manifestou ser favorável a essa ideia, mas questões importantes ainda permanecem sem resposta. Uma é se os usuários do Facebook assumiriam essa tarefa de forma responsável. "Você não pode apenas ter Joe Schmo, que acha que o The New York Times é um antro liberal, [desafiando o jornal] só porque Trump diz que é o 'inimigo do povo', disse **Brooke Binkowski**, ex-editor-gerente da Snopes, um site profissional de verificação de fatos que anteriormente tinha uma parceria com o Facebook. 73 Embora essas hesitações exijam atenção, ainda é possível que o Facebook crie um algoritmo para filtrar alertas baseados em crowdsourcing que diminuam a desinformação das eleições — e faça isso a tempo das eleições em novembro de 2020

Alterações Voltadas às Eleições

Salas de Guerra e Equipes Especiais

Questionado sobre as ações preventivas para as eleições, numa entrevista de abril de 2019, Zuckerberg disse:

"Estou confiante de onde estamos agora." Acrescentou que o Facebook possui, "talvez, alguns dos sistemas mais avançados entre empresas e governos do mundo para impedir o tipo de tática que a Rússia e agora outros países também tentaram adotar". 74

Ele está correto, como vimos, que as redes sociais progrediram, mas existe o perigo de que sua confiança leve à complacência.

Facebook, Twitter e YouTube criaram equipes dedicadas para impedir a manipulação de eleições. A "equipe analítica multifuncional" do Twitter emprega software específico para identificar "atividades anômalas e potencialmente maliciosas, automatizadas e coordenadas por humanos" — por exemplo, um grande número de robôs twittando simultaneamente sobre o mesmo assunto. 75 O "grupo de análise de ameaças" do Google tem uma função semelhante em nome do YouTube, uma subsidiária da empresa. 76 O Facebook possui uma equipe de 500 pessoas dedicado à proteção da integridade das eleições. 77

O Facebook decidiu criar "salas de guerra" nas semanas anteriores às eleições, incluindo as da Índia e da União Europeia em 2019. Essas ações temporárias reúnem dezenas de funcionários de diferentes departamentos da empresa que se concentram intensamente em avaliar comportamento ou conteúdo suspeito. A sala de guerra do Facebook recebe informações da equipe de integridade eleitoral, que, por sua vez, pode contar com cerca de 15.000 funcionários do Facebook e prestadores de serviço que revisam o conteúdo de forma geral. 78

As salas de guerra podem se converter numa estratégia de mídia das empresas, não passando de uma encenação para tirar uma boa fotografia. Mas o simbolismo pode ser construtivo se houver uma sinalização, interna e externa, que a empresa leva a sério a proteção das eleições.

Relações com o Governo

Desde 2016, as três principais empresas de mídia social também estabeleceram ou aperfeiçoaram parcerias com agências governamentais para preservar as eleições. Em alguns casos, o Facebook tem desempenhado o papel de coordenador — por exemplo, ao organizar uma reunião envolvendo todo o setor, com a participação de autoridades federais em sua sede em Menlo Park, Califórnia, em maio de 2018. Zuckerberg falou publicamente sobre as indicações que o Facebook recebe das agências de inteligência, que levaram a eliminação específica de contas suspeitas. 79

O Twitter descreveu seu relacionamento com a Força-Tarefa de Influência Estrangeira do FBI e a Força-Tarefa de Segurança Eleitoral do Departamento de Segurança Nacional. No dia das eleições em 2018, o Twitter disse que "praticamente participou" de um centro de operações de Segurança Nacional, envolvendo representantes do FBI e de outras agências envolvidas no processo. 80 Além disso, o Twitter criou um "portal" dedicado em seu site para que as agências eleitorais estaduais alertem a empresa para novos problemas o mais rápido possível. 81

Propaganda Política

Em 2016, os agentes russos da IRA compraram anúncios de propaganda política on-line com o objetivo de dividir a população, geralmente paga com rublos russos. Desde então, as principais plataformas aumentaram a transparência das propagandas relacionadas a eleições e questões nacionais — principalmente exigindo que sejam rotuladas com avisos de "pago por". Facebook, Twitter e Google/YouTube também passaram a exigir que os patrocinadores de anúncios confirmem suas identidades e localização nos EUA. Todos os três criaram bancos de dados pesquisáveis de propaganda política, que permitem a qualquer pessoa buscar padrões nesse tipo de gasto. 82 O Twitter deu um passo além, ao banir qualquer propaganda da RT (anteriormente chamada de Russia Today) e do canal Sputnik, meios de comunicação russos que foram descritos pelas autoridades de inteligência dos EUA como armas de propaganda do regime de Putin. 83

Embora em tese sejam salutares, as medidas de transparência publicitária podem não ser tão potentes quanto as empresas afirmam. Primeiramente, embora o uso da publicidade pelos russos em 2016 tenha sido importante para conseguir um público americano, a maior parte da desinformação não foi divulgada pelos anúncios comprados. Em vez disso, foi disseminado por posts, compartilhamentos e outras atividades conhecidas como "orgânicas", não pagos. Em segundo lugar, as organizações de notícias dos EUA

“

As salas de guerra podem se converter numa estratégia de mídia das empresas, não passando de uma encenação para tirar uma boa fotografia. Mas o simbolismo pode ser construtivo se houver uma sinalização, interna e externa, que a empresa leva a sério a proteção das eleições.

”

avaliaram a exigência de indicação "pago por" do Facebook e descobriram que ele é facilmente manipulado. Para testar, a **Vice News** comprou anúncios falsos indicando todos os 100 senadores americanos como compradores. As compras passaram pelo processo de aprovação da plataforma. 84

Um projeto de lei federal traz um modelo melhor para a transparência de anúncios. Ele é conhecido como **Lei de Anúncios Honestos**, que vai além das medidas voluntárias adotadas pelas empresas. Apresentado por primeira vez no outono de 2017, o projeto exigiria a divulgação de anúncios políticos online de forma comparável aos requisitos já existentes para a mídia tradicional impressa, por rádio e TV. A Câmara, controlada pelos democratas, aprovou o projeto no início de 2019, como parte de um grande pacote de reforma política. Em maio de 2019, um grupo bipartidário de senadores, incluindo Amy Klobuchar (D., Minn.), Mark Warner (D., Va.) e Lindsey Graham (R., S.C.), apresentou novamente o projeto de Lei de Anúncios Honestos no Senado. Mas enfrenta oposição quase certa do líder da maioria republicana no Senado, Mitch McConnell, de Kentucky, que disse que não vai colocar para votação qualquer norma relacionada às eleições. 85

Essas Medidas Funcionaram?

Já se passou tempo suficiente desde o desastre da desinformação das eleições de 2016. Com isso, os especialistas já começam a avaliar se as medidas adotadas pelas plataformas estão funcionando. Vários estudos compararam a situação nos EUA de 2018 com a de 2016, enquanto outras avaliações consideraram a preparação para as eleições parlamentares da União Europeia em maio de 2019.

Estudos dos Estados Unidos

O Instituto para a Internet de Oxford, um braço da Oxford University, publicou uma **pesquisa** em novembro de 2018, mostrando que a desinformação se tornou mais comum desde 2016. No período de 30 dias antes das eleições minoritárias dos EUA, 25% das notícias compartilhadas pelo Facebook e pelo Twitter relacionadas à eleição tinham conteúdo classificado como "lixo", descobriram os pesquisadores de Oxford. Esse resultado representou um aumento de cinco pontos percentuais em relação às eleições de 2016. A equipe de Oxford definiu como "lixo" as notícias intencionalmente "enganosas, falsas ou incorretas, passando-se por notícias reais sobre política, economia ou cultura". A maioria das notícias identificadas como "lixo" teve origem nos próprios EUA, não na Rússia. 86

Mas outro **estudo**, publicado em fevereiro de 2019, chegou a uma conclusão mais otimista. Cientistas políticos de quatro universidades — Princeton, University of Exeter, Washington University at St. Louis e University of Michigan — descobriram que o Facebook teve um papel "muito menor" na condução de usuários a sites de "notícias falsas" em 2018 do que teve em 2016. De maneira mais ampla, disseram os pesquisadores, "o número de americanos que consomem notícias falsas e o consumo total online de notícias falsas diminuiu aproximadamente 75% da campanha de 2016 para a de 2018". 87 O resultado do estudo dessas quatro universidades basicamente coincidiu com o da **pesquisa** realizada por pesquisadores da NYU e da Stanford University, que

descobriram que a interação com "notícias falsas" no Facebook caiu mais de 50% entre o final de 2016 e meados de 2018. Enquanto isso, afirmaram os pesquisadores, a interação no Twitter aumentou nesse mesmo período. Embora não saibam ao certo as razões que levaram a esse resultado, os pesquisadores da NYU e de Stanford escreveram: "O conjunto de mudanças nos algoritmos e de políticas implementadas pelo Facebook após a eleição de [2016] parece ser uma explicação plausível." Ainda assim, os pesquisadores acrescentaram que, mesmo após a redução das atividades envolvendo notícias falsas no Facebook, o nível absoluto de interação permaneceu alto — cerca de 70 milhões de interações por mês. 88

É claro que a diminuição da desinformação entre 2016 e 2018, indicada por esses estudos, não podem prever ao certo o que ocorrerá em 2020, quando teremos uma disputa presidencial muito mais importante, com potencial para atrair uma interferência muito maior de atores estrangeiros e local nos EUA.

Estudos da União Europeia

Apenas algumas semanas depois das eleições da UE em maio de 2019, a Comissão Europeia e o braço de política externa e segurança do bloco publicaram um **relatório** preliminar sobre atividade de organizações ligadas à Rússia e outros atores. Afirma que essas organizações buscaram minar a credibilidade da eleição por meio de desinformação no Facebook, Twitter e YouTube. "Foram encontradas evidências de atividades contínuas e sustentadas de desinformação promovidas por fontes russas com o objetivo de suprimir a participação de eleitores e influenciar as suas preferências", afirmou o relatório. Ao reconhecer um nível mais elevado de vigilância das plataformas sobre anúncios políticos, a Comissão indica que mais de 600 grupos e páginas do Facebook que operam na França, Alemanha, Itália, Reino Unido, Polônia e Espanha espalharam desinformação e discurso de ódio para milhões de usuários. "As plataformas precisam se esforçar mais para combater efetivamente a desinformação", afirmou a Comissão. 89

Em oposição ao pronunciamento pessimista da Comissão Europeia, uma **pesquisa** voltada para as eleições da UE realizada pelo Instituto para Internet de Oxford não foi capaz de chegar a resultados conclusivos. A respeito do Facebook, os pesquisadores de Oxford verificaram que as notícias individuais classificadas como lixo podem "superar e muito até as melhores e mais importantes histórias produzidas por profissionais, atraindo até quatro vezes mais volume de compartilhamentos, curtidas e comentários". Mas, em termos gerais, ao analisar as fontes compartilhadas no Twitter, os pesquisadores chegaram a descobertas mais animadoras: "Menos de 4% das fontes que circulavam no Twitter durante nosso período de coleta de dados eram lixo ou eram provenientes de fontes conhecidas da Rússia, com usuários compartilhando muito mais links ligados a meios de comunicação em geral (34%)." 90 Embora o desempenho de determinadas notícias virais classificadas como lixo continue sendo um problema importante, a baixa prevalência geral de conteúdo falso parece um sinal de esperança.

Resultado

O resultado desses estudos e a avaliação governamental parece ser que algumas medidas de combate à desinformação nas redes sociais podem estar funcionando até certo ponto. No geral, porém, os resultados apontam para a necessidade de cautela. É preciso estar atentos, pois os atores e as táticas de desinformação provavelmente mudarão em 2020, tornando difícil prever o que ocorrerá nesse período. O Facebook, por exemplo, comemorou publicamente o resultado de algumas das pesquisas citadas aqui, pois afirmam que a empresa está "promovendo avanços no combate a notícias falsas". 91 Mas, como observado pela organização literária e de direitos humanos PEN America, neste momento, "é prematuro fazer qualquer autofelicitação". 92

4. Conclusões e Recomendações

As plataformas estão adotando melhor táticas defesa do que em 2016. Mas o parâmetro de 2016 não deveria servir de referência. Este é o momento para essas empresas intensificarem seus esforços e se prepararem para 2020.

A desinformação tem muitas faces. Este relatório oferece previsões fundamentadas sobre o que pode acontecer em 2020, mas os artistas da desinformação provavelmente têm outros truques nas mangas.

Candid Wüest, pesquisador chefe sênior de ameaças da empresa de segurança cibernética Symantec, disse por e-mail que seus colegas e ele passaram a identificar artifícios especificamente projetados para enganar os filtros de IA das plataformas. Esses métodos podem ser combinados no próximo ano com abordagens discutidas na Parte 2 deste relatório. De acordo com Wüest, uma das técnicas consiste em fazer com que as contas de robôs amplifiquem o conteúdo legítimo com potencial para causar divisão na sociedade —digamos, uma reportagem sobre imigração—, que é mais difícil de ser detectado do que a atividade de robôs circulando conteúdo inventado. Outra reviravolta: os robôs reciclam notícias reais produzidas há meses ou anos, com o objetivo de reviver disputas on-line que podem já ter sido resolvidas. Novamente, a ideia é inflamar os usuários partidários, sem ativar os filtros da plataforma.

Para combater essas estratégias que estão em constante mudança, as empresas de mídia social estão se defendendo melhor do que fizeram em 2016. Mas o parâmetro de 2016 não deveria servir de referência. Este é o momento para elas intensificarem seus esforços e se prepararem para 2020.

As recomendações a seguir se destinam a incentivar as boas práticas para beneficiar os usuários e a sociedade como um todo. Combater a desinformação é uma forma de proteger a democracia, cuja sobrevivência depende de uma comunicação autêntica de fatos reais.

A boa notícia para as plataformas é que o combate à desinformação, mesmo considerando os custos de curto prazo com pesquisa e desenvolvimento e com pessoal, deve ajudar a restaurar a reputação das marcas e esfriar o clamor por uma regulação estatal mais rigorosa. Com o tempo, se as empresas enfrentarem seus desafios com mais eficiência, elas colherão bons frutos, da seguinte forma: Os anunciantes pagarão pela atenção dos usuários. Os usuários continuarão a oferecer a atenção deles apenas se confiarem em determinada plataforma. Portanto, a construção de confiança deve ser uma parte central de qualquer plano de negócios das redes sociais. E o combate à desinformação é uma excelente forma de alcançar esse objetivo.

Recomendações para as Empresas de Mídia Social

1 Identifique e Elimine Deepfakes.

Vídeos realistas, mas fraudulentos, podem minar candidatos políticos e agravar o cinismo dos eleitores. De acordo com ex-altos funcionários do setor de política exterior, deepfakes simulando atividades militares ou atos de terrorismo podem até contribuir para desencadear uma guerra. 93 Outro perigo é que o advento da fabricação de vídeos convincentes (e áudio) poderia ofuscar a linha entre a mentira e a realidade de tal forma, que figuras públicas desonestas poderiam exibir conteúdo verdadeiro como deepfakes — um fenômeno que algumas pessoas chamaram de "**dividendo do mentiroso**". 94

Um projeto de lei no Congresso busca responsabilizar criminalmente aqueles que produzem deepfakes sem anunciar as alterações. Uma abordagem melhor, capaz de evitar o risco de configurar censura por parte do governo, seria as redes sociais melhorarem a tecnologia de triagem por AI, aprimorarem a revisão humana e removerem os deepfakes antes de causar muitos danos. Numa iniciativa promissora, o Google tem auxiliado grupos externos de pesquisa que trabalham na detecção de deepfakes. 95 Mais esforços desse tipo serão bem-vindos.

2 Elimine o conteúdo comprovadamente falso de forma geral.

Além dos deepfakes, as empresas de mídia social deveriam eliminar conteúdo comprovadamente falso como um todo. Essa ideia não é tão drástica como possa parecer. As plataformas já identificam e eliminam determinados tipos de conteúdo, desde discursos de ódio até supressão de eleitores. Recomendamos que eles adicionem mais uma categoria: material identificado de forma definitiva como falso. O **Pinterest**, uma rede social focada em receitas e projetos domésticos, iniciou um projeto nesse sentido. De acordo com as diretrizes da comunidade, o Pinterest elimina "conselhos nocivos, conteúdo que busque perseguir pessoas ou grupos protegidos e conteúdo criado como parte de campanhas de desinformação". 96

Não esperamos que o Facebook, Twitter e YouTube cheguem a ficar livres de conteúdo falso. Mas é plausível esperar que a revisão por algoritmo e por pessoas seja capaz de identificar itens potencialmente falsos, e as plataformas façam a avaliação e eliminem o que for comprovadamente falso. Isso pode representar uma redução apenas marginal no volume de conteúdo falso presente nas redes sociais, mas seria um passo importante na direção certa. Como observado anteriormente, conclamamos as empresas a priorizar os conteúdos falsos relacionados a instituições democráticas, começando com as eleições. E sugerimos que elas mantenham um repositório de material removido, acessível ao público e pesquisável, que sirva de base para o estudo por pesquisadores e outros interessados, mas que não possa ser compartilhado nem retuitado. Por fim, quando as plataformas começarem a excluir conteúdo comprovadamente falso, elas precisarão instituir um processo direto de recurso, permitindo que os usuários peçam a revisão e uma reversão, no caso de exclusão equivocada.

3 Contrate um supervisor de conteúdo.

Além de instituir uma política de comprovação de conteúdo falso, cada empresa deve nomear um funcionário graduado para supervisionar o processo de combate à desinformação. Como foi dito em nosso último relatório sobre desinformação produzida localmente, a responsabilidade pelas decisões de conteúdo agora tende a se espalhar entre equipes distintas nas empresas de mídia social. A centralização simplificaria os principais processos e destacaria a sua importância, interna e externamente. É mais provável que isso ocorra se o novo supervisor estiver subordinado diretamente ao COO ou ao CEO.

A escolha das pessoas certas para esses trabalhos confidenciais será fundamental. Uma opção seria contratar um editor ou executivo experiente do setor de jornalismo, que tenha o conjunto certo de conhecimento e bom-senso sobre o que é real e o que não é. Um certo grau de rigidez é outro requisito para o trabalho, já que a responsabilidade do supervisor de conteúdo exigirá uma atuação firme nos casos de alegações infundadas de preconceito apresentadas por ativistas de direita e pelo presidente Trump. Diante do ataque partidário que está ocorrendo, as empresas devem colocar pessoas nesses cargos com uma reputação exemplar, para garantir que todos os usuários sejam tratados de maneira imparcial quando fizerem login.

4 Faça alterações no Instagram.

Em março de 2019, uma manchete no site de notícias tecnológicas **The Verge** alertou: "O Instagram está passando por um momento crucial por conta de informações falsas." O subtítulo do artigo acrescentou: "Postagens anti-vacina e desinformação estão acabando com a plataforma." 97 O artigo da Verge concorda, em linhas gerais, com as conclusões do Comitê de Inteligência do Senado, de que o Instagram gerou mais do dobro de atividade relacionada à desinformação russa em 2016 do que sua empresa mãe, o Facebook.

E o problema não é a falta de tecnologia. Parece ser a falta de uma estratégia clara para lidar com os sérios problemas inerentes ao modelo de funcionamento do Instagram. Atualmente, a plataforma está testando um sistema que usa reconhecimento de imagem e outras ferramentas para encontrar informações potencialmente falsas, que são enviadas para conferência no Facebook. De acordo com a revista **Wired**, o material considerado falso não é recomendado para novos usuários, mas o Instagram não o remove ou diminui sua classificação nos principais feeds dos usuários. O foco da empresa é "ajustar o algoritmo para dificultar a exposição de desinformações para novos usuários, em vez de impedir o alcance das desinformações". 98 Mas os objetivos não são mutuamente exclusivos. Por que não buscar as duas coisas?

O Instagram, com certeza, conseguiu avançar em determinadas áreas. Ficou mais fácil para os usuários identificar contas suspeitas, divulgando informações como o local das contas e os anúncios que estão sendo exibidos. No ano passado, o Instagram também eliminou centenas de contas que exibiram comportamentos não autênticos de forma coordenada. Mas essas mudanças não foram capazes de redimir a reputação crescente da plataforma como veículo de conteúdo falso. O Instagram precisa se mover de forma rápida e agressiva para proteger todos os usuários contra desinformação.

5 Limite o alcance do WhatsApp.

Como um serviço criptografado destinado à comunicação privada, não ao compartilhamento público, o WhatsApp é muito diferente do Instagram, Facebook, Twitter e YouTube. Esse fato aponto para a necessidade de diferentes abordagens para proteger os seus usuários. O WhatsApp está se movendo na direção certa, mas precisa ir além. Originalmente, o serviço de mensagens permitia que os usuários encaminhassem conteúdo para até 256 grupos de bate-papo por vez. Cada grupo pode ter até 256 membros. A multiplicação simples revela que uma pessoa poderia, sozinha, entregar uma mensagem a 65.536 outras. Se as campanhas políticas quisessem que um boato falso se tornasse viral, esses números pareceriam atraentes. Para diminuir esse risco de abuso, o WhatsApp reduziu gradualmente o número de grupos para os quais um usuário pode encaminhar alguma mensagem ao mesmo tempo. O limite atual é de cinco, reduzindo o público máximo de uma mensagem encaminhada para 1.280 destinatários. Essa redução **diminuiu o volume global de encaminhamento** em 25%, de acordo com o Facebook. 99 Para reduzir ainda mais o potencial de abuso, a empresa deveria restringir o encaminhamento para um único grupo de chat, ou 256 pessoas, por vez. Essa medida preservaria a utilidade do WhatsApp como serviço privado de mensagens, tornando mais difícil a sua exploração.

6 Combata ações de desinformação com fins lucrativos.

As empresas de mídia social devem prestar mais atenção ao conteúdo falso distribuído por empresas, consultores e empresas de relações públicas. A desinformação com fins lucrativos não é uma atividade nova. Durante anos, alguns fornecedores de clickbait usaram teorias da conspiração e notícias falsas para atrair usuários ingênuos das redes sociais para seus sites carregados de anúncios. Mas os métodos com fins lucrativos estão evoluindo e se tornando mais ambiciosos do que os meros esquemas de clickbait. Os serviços de desinformação estão à venda em larga escala.

A eliminação pelo Facebook de uma rede vinculada à empresa israelense Archimedes Group revelou uma operação semelhante à IRA da Rússia, sendo que aquela aparentemente operava para ganhar dinheiro, não para alcançar um objetivo ideológico mais amplo. O alcance internacional do Arquimedes —da África à América Latina e ao Sudeste Asiático— serviu para demonstrar que os motivos capitalistas estão transformando a fabricação de conteúdo falso num negócio global. 100

(continua na p. 20)

7 Apoiar a legislação sobre propaganda política e supressão de eleitores.

Propaganda política: As empresas de mídia social devem fazer uso do seu considerável poder de lobby em Washington para conseguir a aprovação da Lei de Anúncios Honestos. Como observado anteriormente, essa medida estenderia as exigências de transparência de propaganda política para anúncios on-line, como ocorre com a mídia impressa, de rádio e televisão. Facebook e Twitter já manifestaram seu apoio à Lei de Anúncios Honestos, mas parece que eles não fizeram da legislação uma prioridade política. A Câmara já aprovou a proposta, mas as empresas devem se empenhar mais para conseguir a sua aprovação no Senado. Até o momento, a liderança republicana no Senado impediu a votação da Lei de Anúncios Honestos, mas o apoio em massa das principais plataformas pode fazer a diferença. Dependendo das eleições de 2020, o bloqueio republicano pode desaparecer. Quanto mais cedo as empresas apoiarem a divulgação on-line obrigatória, melhor.

Supressão de eleitores: Merece destaque o fato de as três principais plataformas terem reforçado suas políticas nessa área, comprometendo-se a remover conteúdo falso destinado a confundir ou intimidar os eleitores no dia da eleição. Mas essas políticas voluntárias não vão longe o suficiente para impedir a supressão de eleitores, que normalmente tenta impedir que negros e latinos, em particular, participem da nossa democracia. Existe um projeto de lei no Congresso, conhecido como **Lei de Práticas Enganosas e Intimidação de Eleitores**, que busca esclarecer que todas as formas de desinformação intencional dos eleitores são proibidas e puníveis com pena de prisão de até cinco anos. Como muitas ações de supressão hoje ocorrem nas redes sociais, as plataformas têm o dever de apoiar essa medida, mesmo que continuem removendo o conteúdo voltado à supressão de eleitores.

8 Melhore o nível de colaboração do setor como um todo.

A cooperação entre as principais plataformas melhorou, mas às vezes elas ainda seguem suas políticas de forma isolada ao lidar com conteúdo prejudicial. Em março de 2019, pesquisadores da **Aliança para Garantia da Democracia** apontaram que, depois que o Facebook eliminou 2.600 páginas, grupos e contas envolvidos em operações coordenadas de informação, "contas relacionadas no Twitter, YouTube e Instagram continuaram operando, espalhando notícias falsas". 101 Essa falta de coordenação representa um desperdício, pois é possível conseguir um avanço muito maior contra as forças da desinformação.

De maneira mais ampla, as redes sociais devem estimular a coordenação, formando uma força-tarefa permanente envolvendo todo o setor, dedicada ao combate à desinformação. As plataformas já trabalharam juntas em várias situações. Existe o Fórum Global da Internet para o Combate ao Terrorismo e a Iniciativa de Rede Global, com foco na liberdade de expressão e privacidade. A Iniciativa PhotoDNA trata de pornografia infantil. E um banco de dados de "rastros digitais" permite que as plataformas eliminem conteúdo extremista violento de forma mais eficiente. Esses esforços conjuntos tiveram graus variados de sucesso, mas giram em torno de um espírito de cooperação que deve alimentar o esforço para limitar a desinformação.

9 Promova a educação em redes sociais de maneira mais direta e sustentada.

Aplaudimos as empresas de mídia social pelo apoio financeiro e incentivo que forneceram a diversas campanhas de educação digital. Esses programas —que variam de simulações em sala de aula para alunos do ensino fundamental e do ensino médio, a vídeos de personalidades populares da Internet— foram concebidos para educar os usuários sobre conteúdo questionável, para que sejam menos propensos a serem enganados ou compartilhados. A partir dessas ações que são dignas de elogio, as empresas devem fazer mais nesse sentido.

Em especial, as plataformas devem tornar as aulas de educação digital um recurso permanente e de destaque disponível em cada um dos seus sites. Sem dúvida, as empresas preferem não lembrar os usuários toda vez que efetuam login que a desinformação é um ponto fraco das redes sociais. Mas essa é a realidade. Quanto mais os usuários são lembrados desse fato —e são ensinados a distinguir o verdadeiro do falso—, menos influência o conteúdo falso exercerá. Campanhas concisas e incisivas de educação, talvez apresentadas no formato de perguntas frequentes (FAQs), deveriam estar disponíveis para todos os usuários de todas as plataformas, de modo permanente.

Notas de Fim

1. Sarah Mervosh, "Distorted Videos of Nancy Pelosi Spread on Face- book and Twitter, Helped by Trump," The New York Times, 24 de maio de 2019 (<https://www.nytimes.com/2019/05/24/us/politics/pelosi-doc-tored-video.html>). Outro vídeo falso amplamente divulgado do tipo "cheapfake", retratou o correspondente da CNN Jim Acosta como sendo fisicamente agressivo contra uma estagiária da Casa Branca durante uma conferência de imprensa presidencial em novembro de 2018. Discutimos o episódio do Jim Acosta no nosso último relatório: Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do," New York University Stern Center for Business and Human Rights, março de 2019 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_domestic_disin-formation_digital?e=31640827/68184927).
2. Glenn Kessler, Salvador Rizzo, and Meg Kelly, "President Trump Has Made 12,019 False or Misleading Claims Over 928 Days," The Washington Post, 12 de agosto de 2019 (<https://www.washingtonpost.com/politics/2019/08/12/president-trump-has-made-false-ormisleading-claims-over-days/>).
3. Clint Watts, "The Kremlin's Strategy for the 2020 U.S. Election: Secure the Base, Split the Opposition," Daily Beast, 29 de maio de 2019 (<https://www.thedailybeast.com/the-kremlins-strategy-for-the-2020-us-election-secure-the-base-split-the-opposition>).
4. James Loke Hale, "More than 500 Hours of Content Are Now Being Uploaded to YouTube Every Minute," Tubefilter, 7 de maio de 2019 (<https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtubeper-minute/>).
5. Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (vide nota 1).
6. Paul M. Barrett, Tara Wadhwa, and Dorothee Baumann-Pauly, "Combating Russian Disinformation: The Case for Stepping Up the Fight Online," New York University Stern Center for Business and Human Rights, julho de 2018 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_domestic_disinformation_digital?e=31640827/68184927).
7. "Full Transcript: Robert Mueller's Statement on the Russia Investigation," Politico, 29 de maio de 2019 (<https://www.politico.com/story/2019/05/29/robert-mueller-statement-russia-investigation-text-transcript-1346453>).
8. Dustin Volz, "No Significant Foreign Interference Seen on Midterm Vote," The Wall Street Journal, 7 de novembro de 2018 (<https://www.wsj.com/articles/u-s-on-alert-for-election-interference-says-nothing-significant-spotted-1541521048>).
9. Ellen Nakashima, "U.S. Cyber Command Operation Disrupted Internet Access of Russian Troll Farm on Day of 2018 Midterms," The Washington Post, 27 de fevereiro de 2019 (https://www.washingtonpost.com/world/nationalsecurity/uscybercommand-operation-disruptedinternetaccessofrussiantrollfactory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9af5bb51b7ff322e9_story.html?utm_term=.85441f897537).
10. Entrevista com o autor.
11. Entrevista com o autor.
12. Eric Schmitt, David E. Sanger, Maggie Haberman, "In Push for 2020 Election Security, Top Official Was Warned: Don't Tell Trump," The New York Times, 24 de abril de 2019 (<https://www.nytimes.com/2019/04/24/us/politics/russia2020election-trump.html>).
13. Daniel R. Coats, "Worldwide Threat Assessment of the U.S. Intelligence Community," Office of the Director of National Intelligence, 29 de janeiro de 2019 (<https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf>).
14. Tom Van de Weghe, "Six Lessons from My Deepfakes Research at Stanford," Medium, 29 de maio de 2019 (<https://medium.com/jsk-class-of-2019/six-lessons-from-my-deepfake-research-at-stanford-1666594a8e50>) and Leo Kelion, "Reddit Bans Deepfake Porn Videos," BBC, 7 de fevereiro de 2018 (<https://www.bbc.com/news/technology-42984127>).
15. John Villasenar, "Deepfakes, Social Media, and the 2020 Election," TechTank, Brookings Institution, 3 de junho de 2019 (<https://www.brookings.edu/blog/techtank/2019/06/03/deepfakes-social-media-and-the-2020-election/>).
16. Samantha Putterman, "Zuckerberg Video about 'Billions of People's Stolen Data' Is a Deepfake," Politifact, 12 de junho de 2019 (<https://www.politifact.com/facebook-fact-checks/statements/2019/jun/12/insta-gram-posts/zuckerberg-video-aboutbillionspeoples-stolen-dat/>).
17. Danielle Keats Citron, "Prepared Written Testimony and Statements for the Record," Hearing on the National Security Challenges of Artificial Intelligence, Manipulated Media, and "Deep Fakes," U.S. House Permanent Select Committee on Intelligence, 13 de junho de 2019 (https://intelligence.house.gov/uploadedfiles/citron_testimony_for_house_committee_on_deep_fakes.pdf).
18. Ali Breland, "The Bizarre and Terrifying Case of the 'Deepfake' Video that Helped Bring an African Nation to the Brink," Mother Jones, 15 de março de 2019 (<https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>).
19. Robert Chesney, Danielle Keats Citron, Quinta Jurecic, "About That Pelosi Video: What to Do About 'Cheapfakes' in 2020," Lawfare, 29 de maio de 2019 (<https://www.lawfareblog.com/about-pelosi-video-what-to-do-about-cheapfakes-2020>).
20. Entrevista com o autor. Vale ressaltar que a mídia tradicional contribui para a divulgação de vídeos distorcidos.
21. Robert S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election," Department of Justice, março de 2019 (<https://assets.documentcloud.org/documents/5955118/The-Mueller-Report.pdf>).
22. "Removing Bad Actors on Facebook," Facebook, 31 de julho de 2018 (<https://newsroom.fb.com/news/2018/07/removingbad-actors-on-facebook/>).
23. Joe Heim et al., "White-Supremacist Rally Near White House Dwarfed by Thousands of Anti-Hate Protesters," The Washington Post, 12 de agosto de 2018 (https://www.washingtonpost.com/local/washington-readies-for-todays-plannedwhitesupremacistrallynearwhitehouse/2018/08/12/551720c49c2811e88d5ec6c594024954_story.html?utm_term=.569627792071).
24. Renee DiResta et al., "The Tactics and Tropes of the Internet Research Agency," New Knowledge for the U.S. Senate Select Committee on Intelligence, 17 de dezembro de 2018 (<https://www.intelligence.senate.gov/press/new-reports-shed-lightinternetresearch-agency%E2%80%9999social-media-tactics>).
25. Stan Schroeder, "U.S. Users Are Leaving Facebook, New Study Claims," Mashable, 7 de março de 2019 (<https://mashable.com/article/facebook-losing-users-us/>).

Endnotes (continued)

26. Alex Pasternack, "It's Not Over: Russia's Divisive Internet Memes Are Still Racking Up Likes," Fast Company, 19 de dezembro de 2018 (<https://www.fastcompany.com/90283167/russia-instagram-war-face-book-memes>).
27. Taylor Lorenz, "Instagram is the Internet's New Home for Hate,"
28. The Atlantic, 21 de março de 2019 (<https://www.theatlantic.com/technology/archive/2019/03/instagram-is-the-internets-new-home-for-hate/585382/>).
Madhumita Murgia, Stephanie Findlay, and Andres Schipani, "India: The WhatsApp Election," Financial Times, 5 de maio de 2019 (<https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaf6d84>); Vinu Goel and Sheera Frenkel, "In India Election, False Posts and Hate Speech Flummox Facebook," The New York Times, 1o de abril de 2019 (<https://www.nytimes.com/2019/04/01/technology/india-elections-facebook.html>), and Philippa Williams and Lipika Kamra, "Technology Could Torpedo India's First WhatsApp Election," Quartz, 4 de março de 2019 (<https://qz.com/india/1563318/indias-2019-election-is-threatened-by-fake-news-on-whatsapp/>).
29. Entrevista com o autor.
30. Alice Revelli and Lee Foster, "Network of Social Media Accounts Impersonates U.S. Political Candidates, Leverages U.S. and Israeli Media in Support of Iranian Interests," FireEye, 28 de maio de 2019 (<https://www.fireeye.com/blog/threat-research/2019/05/social-media-network-im-personates-us-political-candidates-supports-iranian-interests.html>).
31. Nathaniel Gleicher, "Removing More Coordinated Inauthentic Behavior From Iran," Facebook, 28 de maio de 2019 (<https://newsroom.fb.com/news/2019/05/removing-more-cib-from-iran/>).
32. Yoel Roth, Twitter, 28 de maio de 2019 (<https://twitter.com/yoyoel/status/1133448387657736192>). Ao todo, o Twitter supostamente removeu mais de 7.000 contas falsas vinculadas ao Irã durante os primeiros sete meses de 2019. Vide Craig Timberg and Tony Romm, "It's Not Just the Russians Anymore as Iranian and Others Turn Up Disinformation Efforts Ahead of 2020 Vote," The Washington Post, 25 de julho de 2019 (https://www.washingtonpost.com/technology/2019/07/25/its-not-just-russians-anymore-iranians-others-turn-up-disinformation-efforts-ahead-vote/?utm_term=.4408c57407ff).
33. Sheera Frenkel and Nicholas Fandos, "Facebook Identifies New Influence Operations Spanning Globe," The New York Times, 21 de agosto de 2018 (<https://www.nytimes.com/2018/08/21/technology/facebook-political-influence-midterms.html>).
34. Entrevista com o autor.
35. Entrevista com o autor.
36. Veja, por exemplo, comentários de Lisa-Maria Neudert: "Junk News Dominating Coverage of U.S. Midterms on Social Media, New Research Finds," Oxford Internet Institute, 1o de novembro de 2018 (<https://www.oii.ox.ac.uk/news/releases/junk-news-dominating-coverage-of-us-midterms-on-social-media-new-research-finds/>).
37. Cat Zakrzewski, "The Technology 202: Disinformation Spread by Americans Is 'the Hardest Challenge that We Have,' DHS Official Says," The Washington Post, 12 de abril de 2019 (https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/04/12/the-technology-202-disinformation-spread-by-americans-is-the-hard-est-challenge-that-we-havedhs-official-says/5caf9cf91ad2e567949ec-16c/?utm_term=.ec91773e5f).
38. Vide, por exemplo, Vidya Narayanan et al., "Polarization, Partisanship, and Junk News Consumption Over Social Media in the U.S.," Oxford Internet Institute, 6 de fevereiro de 2018 (<https://comprop.oii.ox.ac.uk/research/polarization-partisanship-and-junk-news/>) and Yochai Benkler, Robert Faris, and Hal Roberts, Network Propaganda: Manipulation, Disinformation, and the Radicalization of American Politics, Oxford University Press, (<https://global.oup.com/academic/product/network-propaganda-9780190923631?cc=us&lang=en&>). Veja, também, Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (vide nota 1). Como discutido em "Tackling Domestic Disinformation," boa parte do conteúdo falso tem origem em redes sociais secundárias, como Reddit e Gab, e depois vai parar no Twitter, YouTube e no Facebook.
39. "Candace Owens: Big Tech Trying to Ban Conservatives Ahead of 2020 Election," Breitbart, 12 de junho de 2019 (<https://www.breitbart.com/politics/2019/06/12/watch-candace-owens-big-tech-trying-to-ban-conservatives-ahead-of-2020-election/>).
40. Vide, por exemplo, Kurt Wagner, "Facebook Bans Alex Jones, Milo Yiannopoulos, Other Far-Right Figures," Bloomberg, 2 de maio de 2019 (<https://www.bloomberg.com/news/articles/2019-05-02/facebook-bans-alex-jones-yiannopoulos-other-far-right-figures>).
41. Lauren Feiner, "Trump Says Facebook, Amazon, and Google Were Colluding With Democrats Against Him," CNBC, 10 de junho de 2019 (<https://www.cnbc.com/2019/06/10/trump-sounds-off-on-facebook-huawei-and-5g.html>).
42. Donald J. Trump, Twitter, 11 de julho de 2019 (<https://twitter.com/realDonaldTrump/status/1149279675660869632>).
43. Ben Brody, "Trump Says He'll Summon Companies After Social Media 'Summit,'" Bloomberg, 11 de julho de 2019 (<https://www.bloomberg.com/news/articles/2019-07-11/trump-s-social-media-summit-brings-fringe-voices-to-white-house>).
44. Caroline Kelly, "It Won't Work: Kamala Harris' Campaign Slams Online Attacks on Her," CNN, 30 de junho de 2019 (<https://www.cnn.com/2019/06/29/politics/kamala-harris-responds-donald-trump-jr-race-african-american-black/index.html>).
45. Adam K. Raymond, "Right-Wing Troll Jacob Wohl Allegedly Behind Fake Buttigieg Sexual-Assault Claim," New York, 30 de abril de 2019 (<http://nymag.com/intelligencer/2019/04/right-wing-troll-set-up-fake-buttigieg-sexual-assault-claim.html>).
46. "A Progressive Activist Defends His Deceptive Tactics," WNYC, 11 de janeiro de 2019 (<https://www.wnycstudios.org/story/progressive-activist-defends-his-deceptive-tactics>).
47. Scott Shane and Alan Blinder, "Democrats Faked Online Push to Outlaw Alcohol in Alabama Race," The New York Times, 7 de janeiro, (<https://www.nytimes.com/2019/01/07/us/politics/alabama-senate-facebook-roy-moore.html>).
48. Tony Romm, Elizabeth Dwoskin, and Craig Timberg, "Facebook Is Investigating the Political Pages and Ads of Another Group Backed by Reid Hoffman," The Washington Post, 7 de janeiro de 2019 (https://www.washingtonpost.com/technology/2019/01/07/facebook-is-investigating-political-pages-ads-anothergroupbacked-by-reid-hoffman/?utm_term=.e1f7350447ac).

49. "A Progressive Activist Defends His Deceptive Tactics" (vide nota 46).

50. Entrevista com o autor.

51. Tony Romm and Craig Timberg, "Facebook Suspends Five Accounts, Including That of a Social Media Researcher, for Misleading Tactics in Alabama Elections," The Washington Post, 22 de dezembro de 2018 (https://www.washingtonpost.com/technology/2018/12/22/facebook-suspends-five-accounts-including-social-media-researcher-misleading-tactics-alabama-election/?utm_term=.6fd96ea070dd and Scott Shane and Alan Blinder, "Democrats Faked Online Push to Outlaw Alcohol in Alabama Race" (vide nota 47).

52. Andy Greenberg, "Alphabet-Owned Jigsaw Bought a Russian Troll Campaign as an Experiment," Wired, 12 de junho de 2019 (<https://www.wired.com/story/jigsawrussiadisinformationsocialmediastinalphabet/>).

53. Adam Entous and Ronan Farrow, "Private Mossad for Hire," The New Yorker, 11 de fevereiro de 2019 (<https://www.newyorker.com/magazine/2019/02/18/private-mossad-for-hire>) and Mark Mazzetti, et al., "Rick Gates Sought Online Manipulation Plans from Israeli Intelligence Firm for Trump Campaign," The New York Times, 8 de outubro de 2018 (<https://www.nytimes.com/2018/10/08/us/politics/rick-gates-psy-group-trump.html>).

54. Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior from Israel," Facebook, 16 de maio de 2019 (<https://newsroom.fb.com/news/2019/05/removing-coordinated-inauthentic-behavior-from-israel/>).

55. Isabel Debre and Raphael Satter, "Facebook Busts Israel-based Campaign to Disrupt Elections," Associated Press, 16 de maio de 2019 (<https://www.apnews.com/7d334cb8793f49889be1bbf89f47ae5c>) and Simona Wineglass, "Who Is Behind Israel's Archimedes Group, Banned by Facebook for Election Fakery?" The Times of Israel, 19 de maio de 2019 (<https://www.timesofisrael.com/who-is-behind-israels-archimedes-group-banned-by-facebook-for-election-fakery/>).

56. Robert S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election" (vide nota 21).

57. "Facebook Press Call," Facebook, 26 de abril de 2019 (https://fbnewsroom.us.files.wordpress.com/2019/04/transcript-9549757_final.pdf).

58. Young Mie Kim, "Voter Suppression has Gone Digital," New York University Brennan Center for Justice, 20 de novembro de 2018 (<https://www.brennancenter.org/blog/voter-suppression-has-gone-digital>).

59. Christopher Bing, "Exclusive: Twitter Deletes Over 10,000 Accounts that Sought to Discourage U.S. Voting," Reuters, 2 de novembro de 2018 (https://www.reuters.com/article/us-usa-election-twitter-exclusive/exclusive-twitter-deletes-over-10000-accounts-seeking-to-discourage-votingidUSKCN1N72FA?feedType=RSS&feedName=technologyNews&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+reuters%2FtechnologyNews+%28Reuters+Tech-nology+News%29).

60. Guy Rosen, "An Update on How We Are Doing at Enforcing Our Community Standards," Facebook, 23 de maio de 2019 (<https://newsroom.fb.com/news/2019/05/enforcing-our-community-standards-3/>).

61. Nathaniel Gleicher, "Coordinated Inauthentic Behavior Explained," Facebook, 6 de dezembro de 2018 (<https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>).

62. Entrevista com o autor.

63. Henry Silverman, "The Next Phase in Fighting Misinformation," Facebook, 10 de abril de 2019 (<https://newsroom.fb.com/news/2019/04/tackling-more-false-news-more-quickly/>).

64. "Zuckerberg Says It's 'Really Important' for Facebook to Form Deepfake Video Policy," CBS News, 27 de junho de 2019 (<https://www.cbsnews.com/news/facebook-zuckerberg-says-its-really-important-to-form-policy-on-deepfake-videos/>).

65. Arjun Kharpal, "Facebook Doesn't Want to Be the 'Arbiter of the Truth,' Top Executive Sheryl Sandberg Says Amid Fake News Criticism," CNBC, 24 de abril de 2017 (<https://www.cnbc.com/2017/04/24/facebook-fake-news-sheryl-sanberg.html>).

66. "Cooper Grills Facebook VP for Keeping Pelosi Video Up," CNN, 25 de maio de 2019 (<https://www.cnn.com/videos/tech/2019/05/25/facebook-monika-bickert-pelosi-video-cooper-intv-sot-ac360-vpx.cnn>).

67. Jessica Leinwand, "Expanding Our Policies on Voter Suppression," Facebook, 15 de outubro de 2018 (<https://newsroom.fb.com/news/2018/10/voter-suppression-policies/>) and Sheryl Sandberg, "A Second Update on Our Civil Rights Audit," Facebook, 30 de junho de 2019 (<https://newsroom.fb.com/news/2019/06/second-update-civil-rights-audit/>).

68. Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse," U.S. House Committee on Oversight and Reform, 22 de maio de 2019 (<https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-KaneK-20190522.pdf>) and "Policies and Safety," YouTube, sem data (https://support.google.com/youtube/answer/2801973?hl=en&ref_topic=9282365).

69. "Our Ongoing Work to Tackle Hate," YouTube, 5 de junho de 2019 (<https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html>).

70. "Continuing Our Work to Improve Recommendations on YouTube," YouTube, 25 de janeiro de 2019 (<https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html>).

71. Yoel Roth, "Information Operations on Twitter: Principles, Process, and Disclosure," Twitter, 13 de junho de 2019 (https://blog.twitter.com/en_us/top-ics/company/2019/information-ops-on-twitter.html).

72. Henry Silverman, "The Next Phase in Fighting Misinformation" (vide nota 63).

73. Sam Levin and Julia Carrie Wong, "'He's Learned Nothing': Zuckerberg Floats Crowdsourcing Facebook Fact-Checks," The Guardian, 20 de fevereiro de 2019 (<https://www.theguardian.com/technology/2019/feb/20/facebook-fact-checking-crowdsourced-mark-zuckerberg>). Snopes afirma que encerrou sua parceria com o Facebook devido ao fato de o contrato consumir muito tempo e não ser muito eficiente. Vide, por exemplo, Daniel Funke, "Snopes Pulls Out of its Fact-Checking Partnership with Facebook," Poynter, 10 de fevereiro de 2019 (<https://www.poynter.org/fact-checking/2019/snopes-pulls-out-of-its-fact-checking-partnership-with-facebook/>).

74. "Interview with Facebook CEO Mark Zuckerberg," ABC News, 4 de abril de 2019 (<https://abcnews.go.com/Business/interview-face-book-ceo-mark-zuckerberg-transcript/story?id=62152829>).

75. Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse" (vide nota 68).

Notas de Fim (continuação)

76. Richard Salgado, "Securing U.S. Election Infrastructure and Protecting Political Discourse," U.S. House Committee on Oversight and Reform, 22 de maio de 2019 (<https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-Sal-gadoR-20190522.pdf>).
77. Katie Harbath and Samidh Chakrabarti, "Expanding Our Efforts to Protect Elections in 2019," Facebook, 28 de janeiro de 2019 (<https://news-room.fb.com/news/2019/01/elections-2019/>).
78. Samidh Chakrabarti, "Fighting Election Interference in Real Time," Facebook, 18 de outubro de 2018 (<https://newsroom.fb.com/news/2018/10/war-room/>).
79. Sheera Frenkel and Matthew Rosenberg, "Top Tech Companies Met with Intelligence Officials to Discuss Midterms," The New York Times, 25 de junho de 2018 (<https://www.nytimes.com/2018/06/25/technology/tech-meeting-midterm-elections.html>) e "Interview with Facebook CEO Mark Zuckerberg" (vide nota 74).
80. Kevin Kane, "Securing U.S. Election Infrastructure and Protecting Political Discourse" (vide nota 68).
81. Del Harvey and Yoel Roth, "An Update on Our Elections Integrity Work," Twitter, 10 de outubro de 2018 (https://blog.twitter.com/en_us/topics/company/2018/an-update-on-our-elections-integrity-work.html).
82. Vide, por exemplo, Del Harvey and Bruce Falck, "Announcing New U.S. Issue Ads Policy," Twitter, 30 de agosto de 2018 (https://blog.twitter.com/en_us/topics/company/2018/Announcing-new-US-issue-ads-policy.html). Pesquisadores relatam problemas técnicos com o banco de dados de publicidade do Facebook: Vide, Matthew Rosenberg, "Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised," The New York Times, 25 de julho de 2019 (<https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html>).
83. "Announcement: RT and Sputnik Advertising," Twitter, 26 de outubro de 2017 (https://blog.twitter.com/en_us/topics/company/2017/Announcement-RT-and-Sputnik-Advertising.html).
84. William Turton, "We Posed as 100 Senators to Run Ads on Facebook. Facebook Approved Them All," Vice News, 30 de outubro de 2018 (https://news.vice.com/en_us/article/xw9n3q/we-posed-as-100-senators-to-run-ads-on-facebook-facebook-approved-all-of-them).
85. Jordain Carney, "Democrats Detail New Strategy to Pressure McConnell on Election Security Bills," The Hill, 18 de junho de 2019 (<https://thehill.com/blogs/floor-action/senate/449174-democrats-detail-new-strategy-to-pressure-mcconnell-on-election>) e Maggie Miller, "Bipartisan Group of Senators Seeks to Increase Transparency of Online Political Ads," The Hill, 8 de maio de 2019 (<https://thehill.com/homenews/senate/442598-bipartisan-group-of-senators-seek-to-increase-transparency-of-online>).
86. Nahema Marchal, et al., "Polarization, Partisanship, and Junk News Consumption on Social Media During the 2018 U.S. Midterm Elections," Oxford Internet Institute, 10 de novembro de 2018 (http://blogs.oii.ox.ac.uk/comprop/wp-content/uploads/sites/93/2018/11/marchal_et_al.pdf).
87. Andrew Guess, et al., "Fake News, Facebook Ads, and Misperceptions," University of Michigan, fevereiro de 2019 (<http://www-personal.umich.edu/~bnyhan/fake-news-2018.pdf>).
88. Hunt Allcott, Matthew Gentzkow, and Chuan Yu, "Trends in the Diffusion of Misinformation on Social Media," Stanford University, outubro de 2018 (<https://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf>).
89. "Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions," European Commission, 14 de junho de 2019 (https://eeas.europa.eu/sites/eeas/files/joint_report_on_disinformation.pdf).
90. Nahema Marchal, et al., "Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook," Oxford Internet Institute, 21 de maio de 2019 (<https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Data-Memo.pdf>).
91. "New Research Shows Facebook Making Strides Against False News," Facebook, 7 de fevereiro de 2019 (<https://newsroom.fb.com/news/2018/10/inside-feed-michigan-lemonde/>).
92. "Truth on the Ballot: Fraudulent News, the Midterm Elections, and Prospects for 2020," PEN America, 13 de março de 2019 (<https://pen.org/wp-content/uploads/2019/03/Truth-on-the-Ballot-report.pdf>).
93. Daniel Benjamin and Steven Simon, "How Fake News Could Lead to a Real War," Politico Magazine, 5 de julho de 2019 (<https://www.politico.com/magazine/story/2019/07/05/fake-news-real-war-227272>).
94. Robert Chesney and Danielle Keats Citron, "Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security," California Law Review, próxima edição de 2019 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954).
95. Daisy Stanton, "Advancing Research on Fake Audio Detection," Google, 31 de janeiro de 2019 (<https://www.blog.google/outreach-initiatives/google-news-initiative/advancing-research-fake-audio-detection/>).
96. "Community Guidelines," Pinterest, sem data (<https://policy.pinterest.com/en/community-guidelines>).
97. Casey Newton, "Instagram Is Facing a Reckoning Over Misinformation," The Verge, 22 de março de 2019 (<https://www.theverge.com/inter-face/2019/3/22/18276479/instagram-reckoning-anti-vaxx-misinformation-conspiracy>).
98. Paris Martineau, "Instagram Can Find Misleading Posts—but Won't Take Them Down," Wired, 8 de maio de 2019 (<https://www.wired.com/story/instagram-find-misleading-posts-wont-take-down/>).
99. "More Changes to Forwarding," WhatsApp, 21 de janeiro de 2019 (<https://blog.whatsapp.com/10000647/More-changes-to-forwarding>).
100. Luiza Bandeira, et al., "Inauthentic Israeli Facebook Assets Target the World," Digital Forensic Research Lab, May 17, 2019 (<https://medium.com/dfirlab/inauthentic-israeli-facebook-assets-target-the-world-281ad7254264>).
101. Jessica Brandt and Bradley Hanlon, "Online Information Operations Cross Platforms. Tech Companies' Responses Should, Too," Lawfare, 26 de abril de 2019 (<https://www.lawfareblog.com/online-information-operations-cross-platforms-tech-companies-responses-should-too>).

Centro de Negócios e Direitos Humanos NYU Stern
Center

Leonard N. Stern School of Business

44 West 4th Street, Suite 800

Nova York, NY 10012

+1 212-998-0261

bhr@stern.nyu.edu

bhr.stern.nyu.edu

© 2019 Centro de Negócios e Direitos Humanos NYU Stern
Center. Todos os direitos reservados. Este trabalho está
protegido sob a Licença Internacional Creative Commons para
Atribuição Não Comercial 4.0. Para visualizar uma cópia desta
licença, acesse <http://creativecommons.org/licenses/by-nc/4.0/>

Traduzido por Klébert Renée Machado Gonçalves
DOCUMENTO ENCAMINHADO SEM REVISÃO EM RAZÃO DA URGÊNCIA DO SOLICITANTE
Serviço de Tradução e Interpretação do Senado – SETRIN/SGIDOC
24 de setembro de 2019

Disinformation and the 2020 Election:

How the Social Media Industry Should Prepare

PAUL M. BARRETT



Contents

Executive Summary	1
1. Introduction	3
2. What to Watch for in 2020	5
3. How Social Media Companies Have Responded to Disinformation...	13
4. Conclusion and Recommendations	17
Endnotes	21

Acknowledgments

This report benefited from the insights of many people, including Kevin Carroll of Wiggin and Dana, Joan Donovan of the Technology and Social Change Research Project at the Harvard Kennedy School, Yaël Eisenstat of the Center for Humane Technology, Lee Foster at FireEye, Dipayan Ghosh of the Platform Accountability Project at the Harvard Kennedy School, Brittan Heller of the Carr Center for Human Rights Policy at the Harvard Kennedy School, Justin Hendrix of the NYC Media Lab, Jeff Jarvis of the Craig Newmark Graduate School of Journalism at City University of New York, Darren Linvill and Patrick Warren of Clemson University, Filippo Menczer of Indiana University, Lisa-Maria Neudert of the Oxford Internet Institute, Nick Pickles of Twitter, Iain Robertson and Andy Carvin of the Atlantic Council's Digital Forensic Research Lab, Laura Rosenberger and Bret Schafer of the Alliance for Securing Democracy, Alexandria Walden and Clement Wolf at Google, and Candid Wüest of Symantec.

We are grateful for financial support from the Craig Newmark Philanthropies, the Open Society Foundations, Jonah Goodhart, and the John S. and James L. Knight Foundation.

Author

Paul M. Barrett is the Deputy Director of the New York University Stern Center for Business and Human Rights.

Executive Summary

“
In terms of sheer volume, domestically generated disinformation now exceeds malign content from foreign sources and will almost certainly be a factor in the next election.
”

This report assesses some of the forms and sources of disinformation likely to play a role during the presidential election campaign in 2020. While midterm Election Day in November 2018 did not feature much Russian interference, there is no guarantee that Russia, and possibly other U.S. antagonists, will refrain from digital meddling in the more consequential 2020 contest. What's more, in terms of sheer volume, domestically generated disinformation now exceeds malign content from foreign sources and will almost certainly be a factor in the next election.

“Deepfake” videos will be deployed to portray candidates saying and doing things they never said or did.

Digital voter suppression will again be one of the main goals of partisan disinformation.

Unwitting Americans could be manipulated into participating in real-world rallies and protests.

For-profit firms based in the U.S. and abroad will be hired to generate disinformation.



Iran and China may join Russia as sources of disinformation.

Domestic disinformation will prove more prevalent than false content from foreign sources.

WhatsApp, the Facebook-owned messaging service, may be misused to provide a vector for false content.

Instagram, more than its parent, Facebook, will be the vehicle of choice for people who wish to disseminate meme-based disinformation.

(Executive Summary continued)

The report explores these risks and analyzes what the major social media companies—Facebook, Twitter, and YouTube (owned by Google)—have done to harden their defenses against disinformation. We also offer recommendations of additional steps the companies need to take to prepare for the almost-certain assault in 2020.

Terminology and Scope

By “disinformation,” we mean a relatively broad category of false or misleading text or imagery which is intentionally or recklessly spread to deceive, propagandize, promote discord, or make money via “clickbait” schemes. For the sake of variety, we sometimes refer to “false content” or “false information.”

We also refer to a narrower category of disinformation—“provably false content”—which we urge the social media companies to remove. To illustrate, consider these hypothetical examples: An article headlined, “The Sandy Hook Massacre Was Staged,” can be proven definitively untrue and ought to be deleted for that reason. By contrast, a story headlined, “Journalists Really Are the Enemy of the People,” may be unsubstantiated and misleading, but it isn’t provably false.

This report doesn’t cover cyberattacks — remote intrusions into computer networks, such as the Russian hack-and-release of Democratic email in 2016. Cyberattacks on campaigns and state voting systems are a distinct danger in 2020, but they are beyond the scope of this report. Also beyond the scope are the privacy and disclosure issues that led Facebook to agree in July 2019 to pay \$5.1 billion in settlements to the federal government and to submit to new types of oversight.

Summary of our Recommendations to Social Media Companies

- 1 Detect and remove deepfake videos.**
Realistic but fraudulent videos have the potential to undermine political candidates and exacerbate voter cynicism.
- 2 Remove provably false content in general.**
The platforms already remove hate speech, voter suppression, and other categories of content; we recommend that they add one more.
- 3 Hire a senior content overseer.**
Each company needs an executive with clout to supervise the process of guarding against disinformation.
- 4 Attend to Instagram’s disinformation problem.**
The photo-posting platform needs the concerted attention of its parent, Facebook.
- 5 Limit the reach of WhatsApp.**
Users should be restricted to forwarding content to one chat group at a time.
- 6 Defend against for-profit disinformation.**
The companies must prepare for false content generated by hired-gun firms.
- 7 Back legislation regulating political ads and punishing voter suppression.** Narrowly tailored bills pending in Congress could help curb some forms of disinformation.
- 8 Improve industry-wide collaboration on disinformation.**
For example, when one platform takes down abusive accounts, others should do the same with affiliated accounts.
- 9 Teach social media literacy in a more direct, sustained way.**
Users have to take responsibility for recognizing false content, but they need more help to do it.

1. Introduction

“
The Nancy Pelosi episode foreshadowed one type of disinformation that could disrupt the 2020 election: deliberately distorted video, amplified via social media.
”

In late May of 2019, two manipulated videos disparaging House Speaker Nancy Pelosi spread across social media. One, posted on a conservative Facebook page called Politics WatchDog, had been altered to slow Pelosi's speech during a public appearance, making it appear as if she drunkenly slurred her words. Rudolph Giuliani, one of President Donald Trump's personal lawyers, shared the video on Twitter, asking, "What is wrong with Nancy Pelosi?" (Giuliani later deleted the tweet.)

At roughly the same time, Trump himself tweeted a separate video of Pelosi—one edited by Fox Business Network. Fox interwove moments from a press conference to make it look like the Speaker continually garbled her words. "PELOSI STAMMERS THROUGH NEWS CONFERENCE," the president declared.¹

The Pelosi episode foreshadowed one type of disinformation that is likely to disrupt the 2020 election: deliberately distorted video, amplified via social media. The most daunting threat comes from deepfakes, which are synthetic videos created with artificial intelligence (AI). Subjects of deepfakes appear to say or do things they never said or did. The crudely distorted Pelosi videos, made with long-available technology, are more appropriately described as "cheafakes." They demonstrate that one doesn't need cutting-edge computer artistry to draw mass attention to a misleading portrayal of a prominent

foe. The "drunken"-speech video alone received more than three million views in a matter of days. Deepfakes that are more difficult to refute could enjoy even wider circulation.

The Pelosi videos suggest another theme of this report: While Russian operatives and other foreign-based actors are all but certain to surface (or resurface) in 2020, a greater volume of disinformation probably will come from domestic U.S. sources. Some of these domestic sources will be obscure websites or social media accounts, such as Politics WatchDog; others will be better known. If he stays true to form, President Trump will use his [Twitter account](#) (62 million followers and counting) to issue statements at odds with the truth. As of early August, he had made more than 12,000 false or misleading claims during his presidency, many of them via Twitter, according to *The Washington Post's* [Fact Checker](#).²

Deepfakes and domestic disinformation are two of our eight predictions for 2020. Here's the full list, in capsule form:

- **Deepfakes** threaten to cloud reality at a time when the existence of objective facts increasingly has been called into question.
- **Domestic disinformation** will come mostly from the political right, but the left will also indulge in social media truth-bending.
- **Unwitting Americans:** The Internet Research Agency or other Russian organizations may try to recruit actual Americans for real-life activities.
- **Instagram**, the Facebook-owned image and video service, may turn out to be more of a disinformation magnet than Facebook, Twitter, or YouTube.
- **WhatsApp**, Facebook's messaging service, could become a troubling source of false content in the U.S., as it has been during elections in Brazil and India.
- **Iran and China** are potential fonts of foreign interference, even as Russia continues to try to rile up the American electorate.
- **For-profit firms** based in the U.S. and abroad will offer their election-manipulation services.
- **Digital voter suppression**, a highly direct means of affecting election outcomes, will once again pose problems on social media.

It's impossible to say with certainty how these stratagems might interact or whether others will come into play as well. What seems safe to expect is that Russia's well-documented meddling in 2016—a multifaceted intrusion that included disguised social media personas, armies of automated bot accounts, and distribution of hacked Democratic emails—will not simply be replicated next year. As disinformation analyst [Clint Watts](#) has put it: "Last cycle's trickery won't work the same in 2020."³

Confusing Policies, Ad Hoc Enforcement

Beyond prognostication, this report examines how social media companies have reacted to politically oriented disinformation. In the case of the Pelosi "drunken"-speech video, there was a notable divergence among the platforms. YouTube promptly removed copies of the altered video. Facebook pointedly did not. Instead, Facebook labeled the video false and reduced its distribution to users. Twitter also declined to pull the video. It's difficult to generalize from this outcome except to say that the platforms' policies are confusing and often seem to be enforced in an ad hoc fashion.

We maintain as a general principle that when provably false content, like the Pelosi video, comes to the attention of social media platforms, they ought to take it down, rather than merely annotating or demoting it. The highest priority should be removing provably false content that affects politics or democratic institutions. Guarding against this abundance of disinformation is a tall order. Five hundred hours of video are uploaded to YouTube every minute.⁴ Given this kind of volume and the velocity with which online content circulates, it would be unreasonable to assume that all disinformation can be removed. But that doesn't mean the companies should do nothing more than what they're already doing. They are best positioned to monitor their sites with algorithmic and human scrutiny, and we believe it's their responsibility to mitigate the damage that disinformation does to our public life.

To preserve a record of provably false content, the platforms each should maintain a searchable but cordoned-off archive of disinformation that they've removed. That way, scholars, journalists, and users generally would be able to view the false content for research or other purposes without its surfacing in ordinary search results, recommendations, or users' feeds. The archive would create a record of digital untruth without contributing to its amplification. And the archive's

contents would provide the basis for appeals by users who believe their post or tweet was removed inappropriately.

Focus on the Companies

This is the fourth report by the NYU Stern Center for Business and Human Rights on online disinformation. Our most recent publication was "[Tackling Domestic Disinformation: What the Social Media Companies Need to Do](#)" (March 2019).⁵ Before that, we published "[Combating Russian Disinformation: The Case for Stepping Up the Fight Online](#)" (July 2018).⁶ We return to the topic because of the pernicious effects disinformation can have on elections. In many cases, it's designed to erode democratic values, heighten cynicism, and exacerbate political polarization.

We continue to focus on what the largest and most influential social media companies ought to do about disinformation. That's because the most obvious alternative—government regulation of online content—would raise immediate free speech concerns about official overreach and censorship. "Governments have only sledgehammers in their tool kits," [Joan Donovan](#), director of the Technology and Social Change Research Project at the Harvard Kennedy School, said in an interview.

The [First Amendment to the U.S. Constitution](#), however, doesn't restrict corporations, which means that social media companies may—and routinely do—moderate content on their platforms. In the absence of government regulation, it is incumbent on the companies to exercise more vigorous self-governance. That means taking a tougher line on disinformation. It is in the companies' enlightened self-interest to do so. Governments in Germany, France, Australia, and other countries lacking an equivalent to the First Amendment are enacting laws to police "hate speech" and "fake news." To reduce the probability of governmental content regulation in the U.S., the social media industry should show that it can close the governance gap when it comes to disinformation.

2. What to Watch for in 2020

“
If the U.S. military
has the ability to take
the Russian Internet
Research Agency offline,
at least for a day or two,
does that mean the 2020
elections are secure?
In a word, no.”

In 2016, Russia's Internet Research Agency (IRA) engaged in what former Special Counsel [Robert S. Mueller III](#) called “multiple systemic efforts to interfere” with the U.S. presidential election.⁷ The IRA has remained active since then, but when the November 2018 midterms rolled around, the amount of Russian meddling was negligible.⁸ It's possible that the IRA decided to keep its powder dry for the 2020 presidential race. It's also possible that more aggressive removal of suspicious accounts by the social media companies hindered the Russians.

And there could have been another factor: [U.S. Cyber Command](#), an arm of the Pentagon, reportedly used its offensive hacking capabilities to temporarily block the IRA from using the internet in early November 2018.⁹

So, does this mean the 2020 elections are secure? In a word, no. “The IRA is a small component of the overall Russian operation,” which also includes Moscow's military intelligence service and possibly others, said [Lee Foster](#), who leads the disinformation team at FireEye, a cybersecurity firm whose clients include some of the major social media companies. “All of these actors rework their approaches and tactics,” he added.¹⁰ As a result, said [Kevin Carroll](#), a former senior counselor at the Department of Homeland Security, the relative calm during election season 2018 doesn't guarantee a repeat performance in 2020.¹¹

President Trump refrains from acknowledging this risk, reportedly because he equates attention to Russian election meddling with skepticism about the legitimacy of his 2016 victory.¹² But the U.S. intelligence community isn't as reticent.

In its 2019 [Worldwide Threat Assessment](#), the Office of the Director of National Intelligence predicted that next year, Russia and other American adversaries “almost certainly will use online influence operations to try to weaken democratic institutions, undermine U.S. alliances and partnerships, and shape policy outcomes.”¹³

Deepfakes

The Worldwide Threat Assessment specifically expressed concern about doctored videos. In 2020, it said, U.S. adversaries “probably will attempt to use deepfake or similar machine-learning technologies to create convincing—but false—image, audio, and video files to augment influence campaigns directed against the United States and our allies and partners.”

The term deepfakes comes from a combination of “deep learning” and “fakes.” It originated in 2017 with an anonymous Reddit user who called himself “deepfakes.” This individual gained attention by using deep-learning algorithms to superimpose faces of celebrities onto the bodies of pornographic actors. Under outside pressure, Reddit banned the individual and porn deepfakes generally, but imitators proliferated elsewhere on the internet.¹⁴

“
‘Deepfakes can
be made by anyone
with a computer, internet
access, and interest in
influencing an election.’
— John Villasenor,
Professor of Electrical
Engineering, Public Policy,
and Management at the
University of California,
Los Angeles
”

The movie industry has altered video footage for decades. But Hollywood’s tricks require skilled technicians, a lot of money, and ample time. Open-source AI democratized video fabrication. “Deepfakes can be made by anyone with a computer, internet access, and interest in influencing an election,” according to [John Villasenor](#), a Professor of Electrical Engineering, Public Policy, and Management at the University of California, Los Angeles.¹⁵

A brief digression for some technical background: The broad term [artificial intelligence](#) (AI) refers to algorithms able to perform humanlike tasks, such as recognizing a human face. Deep learning, a subset of AI, refers to arrangements of algorithms that can learn and make intelligent decisions on their own. Deepfakes are falsified videos made by means of deep learning.

A deep-learning system can produce a persuasive counterfeit by studying photographs and videos of a target person, as well as video of an actor speaking and behaving the way the target will be depicted in the new, bogus video. The technology can, in essence, merge the target and the actor into one. Once a preliminary fake has been produced, a method known as GANs, or generative adversarial networks, makes it more believable. The GANs process seeks to detect flaws in the forgery, leading to improvements addressing the flaws. After multiple rounds of detection and improvement, the deepfake is completed.

On June 7, 2019, two British video artists posted a [deepfake of Mark Zuckerberg](#), Facebook’s founder and CEO, on Instagram. The video had been concocted by an Israeli tech startup called Canny AI. The firm used as its target a video Zuckerberg released publicly in 2017. Canny AI said in an email exchange that it altered the facial movements of the original video to

match those of an actor whose voice was substituted for Zuckerberg’s. The result is visually convincing, but the actor’s voice isn’t a close match for the tech executive’s. Moreover, the artists signaled their intention to satirize Zuckerberg by having him talk like a James Bond super-villain: “Imagine this for a second: one man with total control of billions of people’s stolen data, all their secrets, their lives, their futures.”¹⁶

It’s surprising that more serious deepfakes haven’t already surfaced in American politics. Experts testifying at a House Intelligence Committee hearing in June 2019 agreed that it’s just a matter of time. “Imagine that the night before the 2020 election, a deepfake showed a candidate in a tight race doing something shocking he never did,” suggested [Danielle Keats Citron](#), a law professor at Boston University. “The damage would be irreparable. Elections cannot be undone.”¹⁷ (An alleged deepfake of the president of Gabon, [Ali Bongo](#), delivering a 2019 New Year’s address helped precipitate an unsuccessful coup in the African country.¹⁸)

Citron and two co-authors, writing for the *Lawfare* website, noted that political defamation in the U.S. goes back to Alexander Hamilton’s rivalry with Thomas Jefferson. “What is different today,” they said, “is that the falsehoods involve visual and audio ‘evidence’ that our eyes and ears are deeply inclined to trust (not just written words that might more readily be dismissed).” And courtesy of social media, “the frauds can rapidly reach countless individuals.”¹⁹

Some scholars of disinformation downplay the potential significance of deepfakes, arguing that less-polished frauds—like the Pelosi videos or clumsily Photoshopped still images—could be enough to unsettle an election. “Deepfakes aren’t necessary. A lot of people don’t need ‘real’ evidence,” said [Patrick Warren](#), an economist at

Clemson University who studies social media. In a politically polarized nation, he added, “it’s about ideology and finding things that confirm what [groups of voters] already believe.”²⁰

Warren may have a point when it comes to some committed partisans. But his observation doesn’t negate the more consequential dangers that could result from a series of incriminating and seemingly genuine videos introduced in the run-up to Election Day.

Unwitting Americans

Unlike cutting-edge, AI-driven deep-fakes, some disinformation methods are about the old-fashioned goal of getting people onto the streets. “One thing to look for is more effort to recruit real Americans to unwittingly organize political rallies or create new groups online,” [Laura Rosenberger](#), director of the Alliance for Securing Democracy, said in an interview.

The March 2019 [Mueller report](#) fleshed out our understanding of how IRA operatives, posing as grassroots U.S. activists, mobilized Americans to participate in dozens of rallies in 2016 and thereafter. By expanding on this strategy in 2020, the Russians would accomplish one of their main goals—translating influence online into real-world discord.

In 2016, the technique entailed phony IRA social media personas or groups first announcing events, such as dueling anti-Muslim and pro-Muslim demonstrations that the Russians successfully organized in May 2016 in Houston. The IRA would send direct messages promoting events to their online American followers. From those who indicated interest in attending, the IRA would seek local people to coordinate events. The IRA operative would make an excuse for not attending but would seek out conventional media coverage. Attendance at events varied from a handful of participants to hundreds.²¹

An incident from the summer of 2018 illustrated how this stratagem can unfold. A counterfeit left-wing Facebook account called Resisters, which [Facebook labeled](#) as probably Russian, created an August 10 event to counter a planned white supremacist gathering in Washington, D.C. Resisters, the first Facebook page to promote the counter-protest, coordinated with five other apparently real pages dedicated to opposing the white supremacists, according to Facebook. “These legitimate pages unwittingly helped build interest in [the counter-demonstration] and posted information about transportation, materials, and locations so people could get to the protests,” the company added. Before the demonstrations, however, Facebook determined that Resisters was a fraud and removed it. The company also notified 2,600 users who had indicated interest to Resisters about attending.²²

In the end, only 40 white supremacists showed up for the rally near the White House, and they were greatly outnumbered by counter-protesters.²³ Even though it was unmasked, Resisters accomplished its presumed goal of amplifying division within American society.

Instagram

Instagram hasn’t received as much attention in the disinformation context as Facebook, Twitter, and YouTube, but it played a much bigger role in Russia’s 2016 election manipulation than most people realize. And it could become a crucial Russian instrument next year, according to a report commissioned by the Senate Intelligence Committee.²⁴

Started in 2010, Instagram was acquired by Facebook 18 months later for \$1 billion. Today, Instagram has 1 billion users (compared to 2.38 billion for Facebook, 2 billion for YouTube, and 330 million for Twitter). The photo- and video-posting platform is a common destination for younger users bored by Facebook.²⁵

“

That Instagram has outperformed Facebook as a Russian engagement machine may ‘indicate its strength as a tool in image-centric memetic warfare,’ according to a report commissioned by the Senate Intelligence Committee.

”

Instagram’s image-oriented service makes it an ideal venue for memes, which are photos combined with short, punchy text. Memes, in turn, are a popular vehicle for fake quotes and other disinformation.

In 2016, the IRA enjoyed more user engagement on Instagram than it did on any other social media platform. That was a finding of the Intelligence Committee [report](#) released in December 2018. “Instagram was a significant front in the IRA’s influence operation, something Facebook executives appear to have avoided mentioning in Congressional testimony,” the report said. “Our assessment is that Instagram is likely to be a key battleground on an ongoing basis.” The report was written for the committee by experts at the Tow Center for Digital Journalism at Columbia University and two research firms, New Knowledge and Canfield Research.

Analyzing data from 2015 through 2018, the researchers found there were 187 million user engagements with IRA material on Instagram—more than twice as many as on Facebook (77 million) or Twitter (73 million). One meme posted on Instagram by the IRA’s “Blacktivist” account showed a police officer half-clad in a Ku Klux Klan hood-and-sheet above the statement, “The KKK has infiltrated police departments for years.”²⁶

Many Americans say disinformation is a critical problem that needs to be fixed.

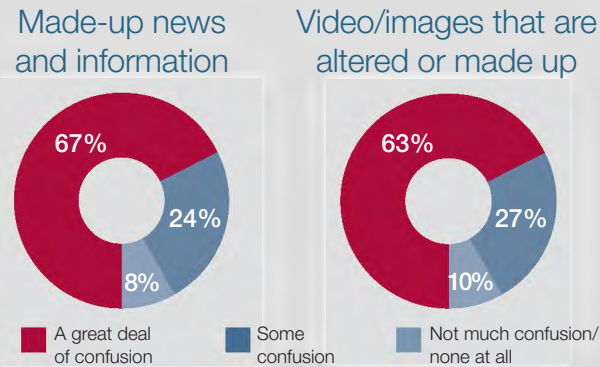
Percentage of Americans who say the issue in question is a very big problem in the country today

70%	Drug addiction
67%	Affordability of healthcare
52%	U.S. political system
51%	Gap between rich and poor
50%	Made-up news/info
49%	Violent crime
46%	Climate change
40%	Racism
38%	Illegal immigration
34%	Terrorism
26%	Sexism

Percentage who say made-up news and information have a very big impact on...

Americans' confidence in government	68%
Americans' confidence in each other	54%
Political leaders' ability to get work done	51%

Two-thirds of U.S. adults think made-up information and altered videos create a great deal of confusion about current issues and events.



Note: Those who didn't answer not shown. Source: Pew Research Center survey of 6,127 U.S. adults, conducted between February 19 and March 4, 2019

Percentage of Americans who have at least "some" trust in social media platform

42%	YouTube	
29%	Facebook	
23%	Instagram	
22%	Twitter	
16%	Snapchat	

Percentage who believe the platform in question is at least "somewhat" responsible for spreading disinformation

64%	Facebook	
55%	Twitter	
48%	YouTube	
46%	Instagram	
39%	Snapchat	

Source: Institute for Public Relations survey of 2,200 Americans between March 19 and March 24, 2019

That Instagram outperformed Facebook as an IRA engagement machine may “indicate its strength as a tool in image-centric memetic warfare,” the committee report concluded. It’s also possible, the researchers acknowledged, that “click farms” created some of the IRA’s Instagram engagement. But that didn’t erase the researchers’ prediction that Instagram would play a major role in 2020.

Other observers have noted that, beyond Russian interference, domestically generated hoaxes and conspiracy theories increasingly are thriving on Instagram. One example is QAnon, a widely circulated right-wing conspiracy theory about a “deep state” plot to take down President Trump. In March 2019, [The Atlantic](#) predicted that Instagram “is likely where the next great battle against misinformation will be fought, and yet it has largely escaped scrutiny.”²⁷ With the approach of the next election, scrutiny will be needed.

WhatsApp

Researchers at the Atlantic Council’s [Digital Forensic Research Lab](#) suggested in an interview that another Facebook property, WhatsApp, deserves a place on our watch list. The case for worrying about WhatsApp rests on two observations: First, the messaging service is encrypted “end-to-end,” meaning that even its corporate overseers can’t see what users are up to. That’s great for privacy but also a recipe for mischief or worse.

Second, WhatsApp has served as a powerful vehicle for disseminating false information during recent presidential elections in both Brazil and India. People in those countries used a WhatsApp forwarding feature to spread wildly misleading content—much of it generated by warring campaigns—to large groups of users. Content on WhatsApp may have more credibility in recipients’ eyes because it often comes from senders with whom recipients are personally familiar. In Brazil, far-right presidential

candidate Jair Bolsonaro triumphed in October 2018, “helped in part by a wave of [toxic rumors and misinformation](#), much of it spread over WhatsApp,” the *Financial Times* reported. Seven months later, Narendra Modi won reelection as India’s prime minister in an election similarly marred by falsehoods and in which WhatsApp became “the central battleground.”²⁸

There are reasons to be skeptical, however, that WhatsApp will be used widely for stirring trouble in the U.S. Acquired by Facebook in 2014 for \$22 billion, the messaging platform has gained tremendous popularity in many parts of the world. It has 1.5 billion users globally, 300 million of them in India and another 120 million in Brazil. But it hasn’t caught on to a similar degree in the U.S., where it has fewer than 60 million adult users. Its smaller U.S. user base limits its potential to spread disinformation. In addition, Americans who use WhatsApp primarily do so for one-to-one or one-to-a-few communication, not one-to-many, as occurred during the Brazilian and Indian elections. With one-to-one messaging, disinformation (or any kind of content) is much less likely to go viral. Finally, WhatsApp has limited the number of groups to which a user can forward a piece of content at one time—a move designed to slow the distribution of untruths.

Still, the Digital Forensic Research Lab cautions that WhatsApp should not be ignored. “It’s been a proven vector [for disinformation] in other countries,” Iain Robertson, the lab’s deputy managing editor, said. “There’s room for that to expand here.”²⁹

Iran and China

Inevitably, other foreign rivals of the U.S. have imitated Russia’s disinformation operation by launching their own. The most likely to surface during the 2020 election is Iran, especially if hostility between Tehran and Washington remains acute.

“

‘Iranians are now following the Kremlin’s playbook,’ Senator Mark Warner (D., Va.) said at the time of an August 2018 removal of 652 fake Facebook accounts, pages, and groups of Iranian and Russian origin.

”

Iran has been one of the most consistent digital foes of the U.S. for the past decade, best known for cyberattacks on banks, hospitals, and government agencies. For the last several years, it has also been test-driving its U.S.-directed disinformation apparatus. In May 2019, FireEye, the cybersecurity firm, released a [report](#) identifying an Iran-based social-media network that used fake American personas espousing both conservative and liberal political views. Some of the phony accounts impersonated real Republican political candidates who ran for office in 2018. The Iranian network promoted anti-Saudi, anti-Israeli, and pro-Palestinian themes. It also expressed support for the 2015 nuclear deal, from which President Trump unilaterally withdrew. Some of the phony personas—using names such as “John Turner” and “Ed Sullivan”—had letters published on these themes in the *New York Daily News*, *Los Angeles Times*, and other mainstream media outlets.³⁰

In response to FireEye’s tip, Facebook removed nearly 100 accounts, pages, and groups from Facebook and Instagram.³¹ Acting independently, Twitter announced that it had removed more than 2,800 “inauthentic accounts originating in Iran.”³² These actions followed two other Facebook takedowns of Iranian accounts in 2018. “Iranians are now following the

“
One of the
domestically generated
conspiracy theories
likely to gain traction
in coming months is
that the major social
media companies
are colluding with
Democrats to defeat
President Trump’s bid
for reelection.”

Kremlin’s playbook,” Senator [Mark Warner](#) (D., Va.) said at the time of an August 2018 removal of 652 fake Facebook accounts, pages, and groups of Iranian and Russian origin.³³

Iran and Russia aren’t the only U.S. adversaries or strategic competitors that are candidates to deploy disinformation in 2020. Enmeshed in a bitter trade war with the U.S., China also could pose a digital threat.

Higher U.S. tariffs or other punitive measures aimed at the Chinese economy could provide sufficient motive for Beijing to unleash “an army of operators” to try to meddle online in 2020, according to [Dipayan Ghosh](#), a former Facebook policy advisor who is co-director of the Platform Accountability Project at the Harvard Kennedy School.³⁴ Kevin Carroll, the former senior counselor at Homeland Security, agreed, saying, “If Sino-U.S. relations further deteriorate, the Chinese could impact U.S. social media simply through brute numbers, with or without the encouragement of the Chinese security services, as so many bilingual and fervently nationalistic Chinese are active online.”³⁵

One major difference between potential Russian and Chinese digital campaigns: The Russians have supported Trump, Ghosh noted, but “the Chinese would probably work to get him out of office because they don’t like his policies or his rhetoric.”

Domestic Disinformation

While foreign election interference has dominated discussion of contemporary disinformation, most purposely false content in the U.S. is generated by domestic sources. That’s the consensus of social media executives, cybersecurity sleuths, and academic researchers.³⁶ Referring to the more plentiful home-grown variety, [Justin Hendrix](#), executive director of NYC Media Lab, said in an interview, “I think we’re going to see a ton of it” in 2020.

[Matt Masterson](#), a senior cybersecurity adviser at the Department of Homeland Security, noted during a recent conference hosted by the Hewlett Foundation that for the federal government, domestically generated disinformation is “the hardest challenge that we have.”³⁷ Foreign interference in a U.S. election is plainly illegal. The U.S. has indicted about two dozen operatives with the Russian IRA and military intelligence service, although it’s highly unlikely Moscow will extradite any of the defendants. By contrast, misleading content created by U.S. citizens can be difficult to distinguish from ordinary political expression, which enjoys First Amendment protection from government regulation or prosecution.

Domestic disinformation emanates from both the political left and right, but as discussed in our March 2019 report, studies show that conservative social media users are more likely to circulate false content.³⁸ One of the domestically generated conspiracy theories likely to gain traction on the right in coming months is that the major social media companies are conspiring with Democrats to defeat President Trump’s bid for reelection. *Breitbart News*, a leading far-right outlet (3.87 million followers on Facebook), has promoted the contention that “big tech companies are trying to make sure conservative voices are suspended ahead of the 2020 election cycle.”³⁹ Over the past year, Facebook, Instagram, Twitter, and YouTube have banned a series of right-wing figures, including Alex Jones, proprietor of the *Infowars* conspiracy website. Explaining these ousters, the social media sites point to their policies against promoting hatred or violence and have denied any vendetta against conservatives.⁴⁰

But President Trump doesn’t buy the denial. In a series of tweets and interviews, he has lashed out at big tech. “I can tell you they discriminate against me,” he told *CNBC*. “You know, people talk about collusion. The real collusion is between the Democrats and these companies.”⁴¹

Introducing a right-wing “social media summit” at the White House in July 2019, Trump tweeted about “the tremendous dishonesty, bias, discrimination, and suppression practiced by certain companies.” Then he threatened: “We will not let them get away with it much longer.”⁴² The president told attendees that he was directing his aides to explore “all regulatory and legislative solutions” for the alleged plot by Facebook, Google, and Twitter.⁴³

The White House gathering seemed intended to rev up conservative disinformation purveyors who, even before Trump gave them his public stamp of approval, had begun to go after would-be Democratic presidential nominees. Some have raised spurious questions about whether Senator [Kamala Harris](#) of California is really an “American black.”⁴⁴ Others have orchestrated false sexual assault allegations against South Bend, Ind., Mayor [Pete Buttigieg](#).⁴⁵

There has also been domestic disinformation from the left. A January 11, 2019, interview on the public radio station WNYC illuminated a liberal social media-distortion scheme from Alabama. “This was a false flag operation,” Democratic activist [Matt Osborne](#) said. “The idea for doing this came from studying what the Russians had done.”⁴⁶

Osborne helped set up a fake “Dry Alabama” Facebook page and companion Twitter account which appeared to have been created by Baptist teetotalers supporting Republican Roy Moore in a special U.S. Senate race in December 2017. The goal was to associate Moore with a statewide alcohol ban and thereby disillusion moderate, business-oriented Republicans. The scheme was carried out on behalf of Democrat Doug Jones, who ultimately won the election but by all accounts didn’t know about the digital skullduggery.⁴⁷ Democrats reportedly executed another false-flag maneuver employing more than a dozen misleading Facebook pages designed to appeal to conservatives nationally in the run-up to 2018 midterms.⁴⁸

Contending that Republicans are using similar deceptive tactics, Osborne said Democrats have to fight fire with fire. In 2020, he predicted, “you’re going to see a movement toward dark money spending on digital campaigns in the closing days of the race.”⁴⁹ For example, he added in an email exchange: “Dissuasion can be as simple as showing, say, a Republican voter an image of a red wave with a triumphal statement that imbues them with a sense of inevitable victory: ‘No need to bother voting. Trump has got it in the bag.’”⁵⁰

For-Profit Firms

Another person who participated in the Alabama disinformation episode was [Jonathon Morgan](#), CEO of New Knowledge, an Austin, Texas-based social media research firm. His role points to an additional danger in 2020: the involvement of private companies or their executives in crafting disinformation.

Once the mainstream media brought the Alabama case to light, Morgan said he’d been acting in his own capacity as a researcher seeking to understand how digital disinformation works, not to affect the outcome of the Jones-Moore Senate race. Facebook nevertheless suspended Morgan’s account and four others run by individuals involved in the episode. The activity in Alabama was funded by Democratic political donors, including Reid Hoffman, the co-founder of LinkedIn, who publicly apologized and said he hadn’t known about the underhanded tactics.⁵¹

Contacted for comment, Morgan said that “it wasn’t our intention to create a misleading page.” His firm, New Knowledge, was mentioned earlier in this report because it helped write an assessment of Russian interference for the Senate Intelligence Committee.

In Russia, disinformation services are sold without excuse or apology. Jigsaw, a corporate think tank affiliated with Google, decided to test the Russian marketplace by retaining a firm called

“

‘This was a false flag operation. The idea for doing this came from studying what the Russians had done.’ — Democratic activist Matt Osborne, describing a liberal social media-distortion plot from Alabama

”

[SEO Tweet](#), which it discovered advertising online. The price for a modest two-week disinformation campaign aimed at a dummy website Jigsaw made especially for the purpose was a mere \$250. The transaction, carried out in the spring of 2018, was an experiment designed to see how influence operations could be bought in Russia, Jigsaw later told *Wired*.⁵² SEO Tweet didn’t respond to a request for comment sent via Twitter.

Israel, home to a robust tech industry, could also be a source of for-profit disinformation. The satiric deepfake of Mark Zuckerberg was engineered by an Israeli start-up. Separately, a February 2019 article in *The New Yorker* described an Israeli intelligence firm called Psy-Group that reportedly specialized in “covertly spreading messages to influence what people believed and how they behaved.” Among the firm’s alleged activities were two digital campaigns in the U.S.: one that undermined American college students who advocated boycotting Israel and a second that tried to tilt a California hospital-board election. In 2016, Psy-Group unsuccessfully pitched its services to the Trump campaign. The company shut down in 2018, and people formerly associated with it didn’t respond to requests for comment.⁵³

Facebook announced in May 2019 that it was taking action against yet another Israeli outfit, a consulting and lobbying firm called [Archimedes Group](#), whose employees allegedly ran political disinformation campaigns in Africa and around the globe. The operatives misrepresented themselves as locals, including local news organizations, and published allegedly leaked information about African politicians, Facebook said in connection with its ban of Archimedes and deletion of 265 accounts, pages, groups, and events. The deceptive activity allegedly focused on Angola, Niger, Nigeria, Senegal, Togo, and Tunisia and also extended to Latin America and Southeast Asia. Archimedes allegedly spent \$812,000 on Facebook advertising paid for in U.S. dollars, Israeli shekels, and Brazilian reals. Facebook apparently didn't figure out whether Archimedes had an overarching agenda, and there is no evidence that the activity was linked to the Israeli government. "We don't want our services to be used to manipulate people," Facebook's head of cybersecurity, Nathaniel Gleicher, wrote in a company blog post.⁵⁴

Efforts to reach Archimedes and individual executives with the firm were unsuccessful. Before its website disappeared in the wake of Facebook's action, the company reportedly described itself online as taking "every advantage available in order to change reality according to our client's wishes."⁵⁵

Digital Voter Suppression

A final category of disinformation almost certain to surface in 2020 is voter suppression. There's nothing new about steering selected voters away from the polls, but the pervasiveness of social media makes this practice particularly pernicious when it goes online.

Recent examples of attempted voter suppression on social media are profuse. As part of their pro-Trump campaign in 2016, Russian IRA operatives tried to persuade African Americans that

there was no point in voting because Democrats didn't care about them.⁵⁶ Facebook has said that in the weeks before the 2018 midterm election, it found and removed 45,000 pieces of voter suppression content.⁵⁷

Researchers at the [University of Wisconsin](#) found several categories of social media suppression in the fall of 2018. Some posts on Twitter tried to deceive opponents of President Trump by providing incorrect information about the day to vote or announcing that "2018 is the first year citizens can vote by TEXT." Other tweets tried to intimidate liberal voters by telling National Rifle Association members and Republicans to bring their guns to the polls. Yet other social media posts attempted to deter African Americans in 2018 by arguing that voting is pointless or that there's no difference between the two major parties. Many of these messages used hashtags like #dontvote and #dontbeavoter.⁵⁸

Some voter suppression efforts invoked more idiosyncratic arguments. Twitter has said it disabled thousands of anonymous voter-suppression bots in 2018, including one batch of more than 10,000 disguised to look like they were from Democrats. Some of the tweets tried to persuade male Democrats to skip the midterms, because otherwise men would drown out women's voices.⁵⁹

Few behaviors strike as directly at the heart of democracy as confusing or bullying people who are entitled to vote. The social media companies will have to remain vigilant on this front.

“

Few behaviors strike as directly at the heart of democracy as confusing or bullying people who are entitled to vote. The social media companies will have to remain vigilant on this front.

”

3. How Social Media Companies Have Responded to Disinformation

“
‘The companies are getting much better at detection and removal of fake accounts.’
— Dipayan Ghosh, co-director of the Harvard Kennedy School’s Platform Accountability Project
”

In a sense, the major social media companies have been preparing for the 2020 election since 2017, when they began to acknowledge the Russian disinformation problem. Since then, these companies have taken a wide range of general steps to clean up their sites and harden their defenses. They also have put in place a number of measures specifically aimed at protecting election integrity. Both categories—general and election-specific—should have a bearing on 2020.

To their credit, the companies are doing more—more communicating with each other, the government, and outside experts; more deleting of fraudulent accounts; more mobilizing of special teams focused on election irregularities. But there’s still much more to do.

General Changes

Removing Sham Accounts

“The companies are getting much better at detection and removal of fake accounts,” Dipayan Ghosh, co-director of Harvard’s Platform Accountability Project, said in an interview. Facebook uses improved AI to delete automated fake accounts by the billions—2.19 billion in just the first three months of 2019. Most of these accounts were blocked within minutes of their creation, preventing them from doing any harm, according to the company.⁶⁰

Over the past two-and-a-half years, Facebook also has announced dozens of smaller, more-targeted takedowns comprising many thousands of accounts and pages which have demonstrated “coordinated inauthentic behavior.”⁶¹ In Facebook’s opinion, the operators of these accounts and pages have worked together to deceive users about

who runs them and what they’re doing. The company stresses that it punishes misleading behavior, not content. But the behavior in question often includes disseminating disinformation. For its part, Twitter has challenged and taken down millions of fake accounts for similar reasons. And YouTube has done so as well, if to a lesser degree.

Compared to the companies’ relative passivity in 2016, these actions demonstrate greater vigor. In addition to honing their AI-screening algorithms, the platforms have hired thousands of additional staff reviewers and outside contractors to hunt for accounts spewing problematic content. But the impressive numbers of fake or deceptive accounts eliminated also indicate the vast supply of such accounts and the near certainty that the companies aren’t catching them all—or perhaps even most of them. “Very many fake accounts are going undetected and can be used for manipulation,” according to [Filippo Menczer](#), a professor of informatics and computer science at Indiana University who studies disinformation. These bot accounts, he added in an email exchange, “can be used to amplify the spread of misinformation, deepfakes, attacks, fear-mongering, voter suppression, or to fake support for candidates.”⁶²

Demoting and Labeling False Content

Intertwined with the problem of fake accounts and inauthentic behavior is the widespread existence of false content that spreads rapidly on social media platforms. What should happen to disinformation once legitimate accounts and pages begin to share it? To deal with falsehoods that can distort the country's political discourse, the platforms have invested in more precise AI, expanded in-house reviewing, and added more third-party fact-checking capacity. Facebook now farms out potentially false stories and images to 53 fact-checking organizations fluent in 42 languages around the world. In particular, the social network has emphasized that it is “using both technology and people to fight the rise in photo and video-based misinformation”—a commitment that reflects growing concern about deepfakes.⁶³

Facebook's Mark Zuckerberg recently said his company is considering banning and removing all deepfakes, or at least the ones it can identify.⁶⁴ But generally, the platforms don't remove content simply because it's been deemed to be false. Instead, they typically reduce the prominence of the untrue material, based on the notion that their executives and employees shouldn't be “arbiters of the truth.”⁶⁵ Facebook, for example, “down-ranks” falsehoods in users' News Feeds by 80%, labels the content as untrue, and provides context on the topic. This is what Facebook did with the doctored video making Nancy Pelosi seem drunk.⁶⁶

The companies don't apply these content-demotion procedures in a consistent way, however. Facebook has said that in certain contexts, it does remove false content. Examples include false information that creates a risk of physical violence or aims to keep voters from the polls or seeks to interfere with the U.S. census.⁶⁷ Twitter and YouTube have said that they, too, will eliminate voter-suppression content.⁶⁸

In June 2019, YouTube announced that it would begin deleting some content based on its deceptiveness. In other words, in certain circumstances, it would act as an arbiter of the truth. Specifically, YouTube said it would take down videos “denying that well-documented violent events, like the Holocaust or the shooting at Sandy Hook Elementary, took place.”⁶⁹ Videos that perpetuate other falsehoods, such as the contention that the Apollo moon landings were staged or that a phony miracle cure will help patients with a serious illness, are no longer recommended to YouTube users, but they remain available on the platform.⁷⁰

No obvious principle underpins these distinctions.



Acknowledging that it needs to find additional ways to identify disinformation, Facebook has launched an initiative to crowdsource fact-checking.



As mentioned earlier, the NYU Stern Center believes that when social media platforms discover provably false content, especially in the political realm, they should not merely reduce its visibility or have a recommendation algorithm cease to tee it up for users. The platform should remove the material altogether, keeping only a well-marked reference copy in an archive of deleted false content. The material in the archive should be publicly accessible to researchers and users generally but should not circulate via feeds or recommendations. Twitter has taken a step in this direction by releasing substantial datasets of tweets, which it has removed after linking the content to disinformation operations backed by Russia or other countries.⁷¹

Crowdsourcing Fact-Checks

Acknowledging that it needs to find additional ways to identify disinformation, Facebook has launched an initiative to crowdsource fact-checking. With 1 billion total posts a day, the volume of false content on the social network far exceeds what in-house reviewers and outside fact-checkers can handle. To deal with this challenge, Facebook says it is looking into how to harness “community-driven approaches to misinformation.” By this, the company means “relying on groups of people who use Facebook to point to journalistic sources that can corroborate or contradict the claims made in potentially false content.”⁷²

CEO Mark Zuckerberg has spoken approvingly about this idea, but significant questions remain. One is whether Facebook users would handle the task responsibly. “You can't just have Joe Schmo, who thinks *The New York Times* is a liberal rag, [challenging the newspaper] just because Trump says it's the ‘enemy of the people,’” said [Brooke Binkowski](#), the former managing editor of *Snopes*, a professional fact-checking site that previously partnered with Facebook.⁷³ While such hesitations demand attention, it's still possible that Facebook could devise an algorithmic method to filter crowdsourced alerts that would diminish election disinformation—and do it in time for November 2020 voting.

Election-Related Changes

War Rooms and Special Teams

Asked about election defenses in an April 2019 interview, Zuckerberg said: “I'm confident in where we are right now.” Facebook, he added, has “probably some of the most-advanced systems of any company or government in the world for preventing the kind of tactics that Russia and now other countries, as well, have tried.”⁷⁴

He's correct, as we've seen, that the social media platforms have made progress, but there's a danger that his confidence will lead to complacency.

Facebook, Twitter, and YouTube all have established dedicated teams to prevent manipulation of elections. Twitter's "cross-functional analytical team" uses specially designed software to identify "anomalous and potentially malicious automated and human-coordinated activity"—for example, large numbers of bots tweeting simultaneously about the same topic.⁷⁵ Google's "threat analysis group" serves a similar function on behalf of the company's YouTube subsidiary.⁷⁶ Facebook has a 500-member team devoted to protecting election integrity.⁷⁷

Facebook has taken to setting up "war rooms" in the weeks preceding elections, including those in India and the European Union in 2019. These arrangements temporarily bring together several dozen staff members from different departments of the company who focus intensively on evidence of suspicious behavior or content. The Facebook war room gets input from the election-integrity team, which, in turn, can draw on some 15,000 Facebook employees and outside contractors who generally review content.⁷⁸

War rooms are susceptible to being lampooned as mere backdrops for photo ops. But the symbolism can be constructive if it signals, both internally and externally, that the company is serious about protecting elections.

Government Relationships

Since 2016, the three major social media companies also have established or improved election-related partnerships with government agencies. Facebook at times has played a coordinating role—for example, hosting an industry-wide [meeting](#) with federal officials at its Menlo Park, Calif., headquarters in May 2018. Zuckerberg has spoken publicly about

tips Facebook receives from intelligence agencies, which have led to targeted take-downs of suspicious accounts.⁷⁹

Twitter has described relationships with the FBI's Foreign Influence Task Force and the Department of Homeland Security's Election Security Task Force. On Election Day in 2018, Twitter said that it "virtually participated" in a Homeland Security operations center that brought together representatives of the FBI and other relevant agencies.⁸⁰ In addition, Twitter has set up a dedicated "portal" on its site for state election agencies to alert the company to emerging problems as quickly as possible.⁸¹

Political Advertisements

In 2016, Russian IRA operatives bought divisive online political advertisements, often paying in Russian rubles. Since then, the major platforms have increased the transparency of ads related to elections and national issues—chiefly by requiring that they be labeled with "paid for by" notices. Facebook, Twitter and Google/YouTube now also require that ad purchasers confirm their identities and locations within the U.S. All three have created searchable political-advertising databases, which allow anyone to look for patterns in this kind of spending.⁸² Twitter has taken the extra step of banning all advertising by RT (formerly Russia Today) and Sputnik, Russian media outlets that have been described by U.S. intelligence authorities as propaganda arms of the Putin regime.⁸³

Although salutary in theory, the advertising-transparency measures may not be as potent as the companies claim. First, while the Russians' use of advertising in 2016 was important to build U.S. audiences, the majority of disinformation was not spread by purchased ads. Instead, it was disseminated by unpaid-for "organic" activity—posts, shares, and the like. Second, U.S. news organizations have

“

War rooms are susceptible to being lampooned as mere backdrops for photo ops. But the symbolism can be constructive if it signals, both internally and externally, that the company is serious about protecting elections.

”

scrutinized Facebook's "paid for by" notice requirement and found that it is easily manipulated. As a test, [Vice News](#) bought fake ads naming all 100 sitting U.S. senators as the buyers. The purchases sailed through the platform's approval process.⁸⁴

A better model for ad transparency is provided in proposed federal legislation known as the [Honest Ads Act](#), which would go further than the policies the companies have voluntarily adopted. First introduced in the fall of 2017, the bill would mandate disclosure for political ads online in a fashion comparable to already-existing requirements for traditional broadcast and print media. The Democrat-controlled House approved the bill in early 2019 as part of an omnibus political-reform package. In May 2019, a bipartisan group of senators, including Amy Klobuchar (D., Minn.), Mark Warner (D., Va.), and Lindsey Graham (R., S.C.), reintroduced the Honest Ads Act in the Senate. But it faces almost certain opposition from Republican Senate Majority Leader Mitch McConnell of Kentucky, who has said that he is not going to bring election-related legislation to the floor for a vote.⁸⁵

Have These Steps Worked?

Enough time has passed since the 2016 election-disinformation debacle to allow scholars to begin to assess whether the steps the platforms have been taking are working. Several studies have compared the situation in the U.S. in 2018 to that in 2016, while other evaluations have considered the run-up to the May 2019 European Union Parliamentary election.

Studies of the United States

The Oxford Internet Institute, an arm of Oxford University, published [research](#) in November 2018 showing that disinformation had become more common since 2016. During a 30-day period before the U.S. midterms, 25% of Facebook and Twitter shares related to the election contained “junk news,” the Oxford researchers found. That represented an increase of five percentage points from the 2016 election season. The Oxford team defined junk news as deliberately “misleading, deceptive, or incorrect information purporting to be real news about politics, economics, or culture.” Most of the junk news they identified came from domestic U.S. sources, not Russia.⁸⁶

But another [study](#), published in February 2019, reached a more sanguine conclusion. Political scientists from four universities—Princeton, the University of Exeter, Washington University at St. Louis, and the University of Michigan—found that Facebook played a “dramatically reduced” role in leading users to “fake news” sites in 2018 than it had in 2016. More broadly, the researchers said, “the proportion of Americans consuming fake news and their total consumption of fake news websites declined by approximately 75% between the 2016 and 2018 campaigns.”⁸⁷

The four-universities study was roughly consistent with [research](#) by scholars from NYU and Stanford, who found that engagement with “fake news” on Facebook fell by more than 50% from

the end of 2016 to mid-2018. Meanwhile, the researchers said, engagement on Twitter rose during the same period. While tentative about the reasons for this outcome, the NYU and Stanford researchers wrote: “The suite of policy and algorithmic changes made by Facebook following the [2016] election seems like a plausible candidate.” Still, the scholars added, even after the reduction in Facebook fake news engagement, the absolute level of interaction remained high—roughly 70 million interactions per month.⁸⁸

Studies showing a decrease in disinformation between 2016 and 2018 cannot, of course, predict with any certainty what might happen in 2020, when a much more momentous presidential contest is likely to attract intensified interference by both foreign and domestic U.S. actors.

Studies of the European Union

Just weeks after the E.U. elections in May 2019, the European Commission and the bloc’s foreign policy and security arm published a preliminary [review](#) that found that Russian-linked organizations, among other actors, had sought to undermine the credibility of the election by means of disinformation on Facebook, Twitter, and YouTube. “The evidence collected revealed continued and sustained disinformation activity by Russian sources aiming to suppress turnout and influence voter preferences,” the review said. While praising the platforms’ greater vigilance about political ads, the Commission noted that more than 600 Facebook groups and pages operating across France, Germany, Italy, the United Kingdom, Poland, and Spain reportedly spread disinformation and hate speech to millions of users. “More needs to be done by the platforms to effectively tackle disinformation,” the Commission said.⁸⁹

In contrast to the European Commission’s gloomy pronouncement, [research](#) on the E.U. elections by the Oxford Internet Institute produced mixed results. On Facebook, the Oxford researchers found, individual junk news stories can “hugely outperform even the best, most important professionally produced stories, drawing as much as four times the volume of shares, likes, and comments.” But looking generally at sources shared on Twitter, the researchers described more heartening findings: “Less than 4% of sources circulating on Twitter during our data collection period were junk news or known Russian sources, with users sharing far more links to mainstream news outlets overall (34%).”⁹⁰ Although the performance of particular viral junk news stories remains a significant problem, the low overall prevalence of false content seems like a hopeful sign.

The Upshot

The upshot of this scholarship and governmental evaluation appears to be that some social media anti-disinformation measures may be working to some extent. Overall, though, the results argue for caution. And it bears repeating that disinformation actors and tactics will likely shift in 2020, meaning that all bets will be off.

Facebook, for one, has publicly celebrated some of the research cited here, which it claims shows that the company is “making strides against false news.”⁹¹ But as the literary and human rights organization [PEN America](#) has observed, at this point, “any self-congratulation is premature.”⁹²

4. Conclusions and Recommendations

“
The platforms are playing better defense than they did in 2016. But that’s an exceedingly low bar. Now is the time for these companies to step up their games in anticipation of 2020.”

Disinformation has many permutations. This report offers educated guesses about some of what may unfold in 2020, but disinformation artists likely have other tricks up their sleeves.

[Candid Wüest](#), a senior principal threat researcher at the cybersecurity firm Symantec, said via email that he and his colleagues are seeing ruses specifically designed to fool platform AI filters. These methods could be combined next year with approaches discussed in Part 2 of this report. One technique, according to Wüest, is to have bot accounts amplify legitimate content that’s divisive—a news report on the immigration controversy, say—which is more difficult to detect than bots trumpeting made-up material. Another twist: bots recycle real news articles that are months or years old, with the aim of reviving online disputes that may have settled down. Again, the idea is to inflame partisan users without activating platform filters.

Confronting these ever-changing strategies, the social media companies are playing better defense than they did in 2016. But that’s an exceedingly low bar. Now is the time for them to step up their games in anticipation of 2020.

The recommendations that follow are intended as practical-minded encouragement to do the right thing for the sake of users and society at large. Protecting against disinformation is a way of protecting democracy, which depends for its survival on authentic communication of real facts.

The good news for the platforms is that fighting disinformation, even if there are short-term R&D and personnel costs, ultimately should help restore their brand reputations and slow demands for draconian government regulation. Over time, if the companies address their challenges more effectively, they will see a benefit to the bottom line, as follows: Advertisers pay for users’ attention. Users will continue to offer their attention only if they trust a given platform. Building trust, therefore, should be a central part of any social media business plan. And countering disinformation offers an excellent way to do it.

Recommendations to the Social Media Companies

1 Detect and remove deepfakes.

Realistic but fraudulent videos have the potential to undermine political candidates and exacerbate voter cynicism. Former senior foreign policy officials have warned that deepfakes simulating military activity or acts of terrorism could even help spark a war.⁹³ Another danger is that the advent of convincing video (and audio) fabrication could so blur the line between sham and reality that dishonest public figures would successfully wave off genuine depictions as deepfakes—a phenomenon that some have called the “[liar’s dividend](#).”⁹⁴

Legislation has been introduced in Congress that would criminally punish those who make deepfakes without disclosing the modifications. A better approach, and one that avoids the danger of overreaching government censorship, would be for the social media platforms to improve their AI-screening technology, enhance human review, and remove deepfakes before they can do much damage. In one promising development, Google has been assisting outside research groups working on deepfake detectors.⁹⁵ More such efforts are necessary.

2 Remove provably false content in general.

Beyond deepfakes, the social media companies should take down provably false content in general. This idea isn’t as drastic as it may sound. The platforms already identify and remove certain categories of content, ranging from hate speech to voter suppression. We’re recommending that they add one more category: material that can be definitively shown to be untrue. [Pinterest](#), a social network that focuses on recipes and home projects, has blazed a path in this regard. According to its community guidelines, Pinterest removes “harmful advice, content that targets individuals or protected groups, and content created as part of disinformation campaigns.”⁹⁶

We don’t expect Facebook, Twitter, and YouTube ever to be swept clean of falsehoods. Instead, as algorithmic and human review flag potentially untrue items, the platforms would evaluate them and purge what’s provably false. This may put only a modest additional dent in the volume of social media fakery, but it would be an important step in the right direction. As noted earlier, we urge the companies to prioritize false content related to democratic institutions, starting with elections. And we suggest that they retain clearly marked copies of removed material in a publicly accessible, searchable archive, where false content can be studied by scholars and others, but not shared or retweeted. Finally, as the platforms begin to delete provably false content, they will need to provide a straightforward appeals process allowing users to seek timely review and a remedy when mistakes are made.

3 Hire a content overseer.

Hand-in-hand with instituting a policy on provably false content goes the appointment at each company of a senior official to oversee the process of guarding against disinformation. As we argued in our last report on domestically generated disinformation, responsibility for content decisions now tends to be scattered among disparate teams within the social media companies. Centralization would streamline key processes and underscore their importance, internally and externally. This is more likely to happen if the new official reports directly to the COO or CEO.

Choosing the right people for these sensitive jobs will be critical. One option is hiring an experienced editor or executive drawn from the journalism business who has the right combination of savvy and commonsense about what’s real and what’s not. A stiff spine is another job requirement, as the content overseer’s responsibility will include standing firm in the face of unfounded claims of bias by right-wing activists and President Trump. In light of the ongoing partisan onslaught, the companies should put people in these posts who are beyond (reasonable) reproach and who will make sure all users are treated even-handedly every time they log in.

4 Make changes at Instagram.

In March 2019, a headline on the tech news site [The Verge](#) warned: “Instagram is facing a reckoning over misinformation.” The article’s subhead added: “Anti-vaccine posts and misinformation are rotting the platform.”⁹⁷ The message of the *Verge* piece combines ominously with the finding in the Senate Intelligence Committee report that Instagram generated more than twice as much engagement with Russian disinformation in 2016 as its parent, Facebook.

The problem isn’t a lack of technology. It appears to be a lack of a clear strategy for addressing the serious problems inherent in Instagram’s operating model. The platform currently is testing a system that uses image recognition and other tools to find potential misinformation, which is sent to Facebook fact-checkers. According to [Wired](#) magazine, material deemed false isn’t recommended to new users, but Instagram doesn’t remove or down-rank it in users’ main feeds. The company chooses to focus “on making it harder for new users to be algorithmically exposed to misinformation, rather than stemming the reach of misinformation.”⁹⁸ But the goals aren’t mutually exclusive. Why not pursue both?

Instagram, to be sure, has made progress in certain areas. It has made it easier for users to identify suspicious accounts by disclosing such information as accounts’ location and the ads they’re running. In the past year, Instagram also has removed hundreds of accounts for displaying coordinated inauthentic behavior. But these steps haven’t cured the platform’s burgeoning reputation as a vehicle for false content. Instagram needs to move swiftly and aggressively to protect all of its users from disinformation.

5 Limit the reach of WhatsApp.

As an encrypted service meant for private communication, not public sharing, WhatsApp is very different from Instagram, Facebook, Twitter, and YouTube. That means different approaches are necessary to protect its users. WhatsApp has been moving in the right direction and now needs to go further.

Originally, the messaging service allowed users to forward content to up to 256 chat groups at a time. Each group could have up to 256 members. Simple multiplication reveals that one person could single-handedly deliver a message to 65,536 others. If political campaigns wanted a fake rumor to go viral, these numbers would look attractive. Seeking to diminish this potential for abuse, WhatsApp has gradually reduced the number of groups a user may forward to at one time. The current limit is five, which means that the maximum audience for one forwarded message has shrunk to 1,280. This reduction has [lessened forwarding](#) by 25% globally, according to Facebook.⁹⁹ To further reduce the potential for misuse, the company should restrict users to forwarding to one chat group, or 256 people, at a time. This would preserve WhatsApp’s usefulness as a private messaging service while making it more difficult to exploit.

6 Defend against for-profit disinformation.

The social media companies must pay more attention to false content distributed by corporations, consultants, and public relations firms. Profit-driven disinformation isn’t brand new. For years, some clickbait purveyors have used conspiracy theories and phony news to draw gullible social media users to their ad-laden websites. But for-profit methods are evolving and becoming more ambitious than mere clickbait schemes. Disinformation services are for sale on a large scale.

Facebook’s shutdown of a network linked to the Israeli firm Archimedes Group revealed an operation similar to the Russian IRA but one apparently operated to make money, not accomplish an overarching ideological goal. The international reach of Archimedes—from Africa to Latin America to Southeast Asia—served notice that capitalist motives are turning the manufacture of false content into a global business.¹⁰⁰

(continued on p. 20)

7 Support legislation on political ads and voter suppression.

Political ads: The social media companies should throw their considerable Washington lobbying clout behind the Honest Ads Act. As noted earlier, this measure would extend political-ad disclosure standards to online ads that are comparable to those imposed on traditional print and broadcast media. Facebook and Twitter have endorsed the Honest Ads Act, but they do not seem to have made adoption of the legislation a top political priority. The House has approved the provision, and now the companies should do more to push for its approval in the Senate. To date, the Republican leadership in the Senate has prevented a vote on the Honest Ads Act, but active support by the major platforms could make a difference. Depending on the 2020 elections, the Republican roadblock could disappear. The sooner the companies put their shoulder behind mandatory online disclosure, the better.

Voter suppression: The three main platforms, to their credit, have strengthened policies in this area and vow to remove false content meant to confuse or intimidate voters on Election Day. But these voluntary policies don't go far enough to deter voter suppression, which typically has sought to keep Blacks and Latinos in particular from participating in our democracy. Legislation pending in Congress, known as the [Deceptive Practices and Voter Intimidation Act](#), would clarify that all forms of intentional voter disinformation are prohibited and punishable by up to five years in prison. Given that much modern-day suppression takes place via social media, the platforms have a duty to support this measure, even as they continue to remove voter-suppression content themselves.

8 Improve industry-wide collaboration.

Cooperation among the major platforms has improved, but they still sometimes go their own way when addressing harmful content. In March 2019, researchers at the [Alliance for Securing Democracy](#) pointed out that after Facebook took down 2,600 pages, groups, and accounts engaged in coordinated information operations, “related accounts on Twitter, YouTube, and Instagram continued to operate, spreading falsehoods.”¹⁰¹ This lack of coordination represents a wasted opportunity to make greater progress against the forces of disinformation.

More broadly, the social media industry should foster coordination by forming a permanent inter-company task force devoted to fighting disinformation. The platforms have worked together in a variety of contexts. There's the Global Internet Forum to Counter Terrorism and the Global Network Initiative, which focuses on freedom of expression and privacy. The PhotoDNA Initiative addresses child pornography. And a database of “digital fingerprints” allows the platforms to take down violent extremist content more efficiently. These joint efforts have enjoyed varying degrees of success, but they share a spirit of cooperation that ought to infuse the push to limit disinformation.

9 Teach social media literacy in a more direct, sustained way.

We applaud the social media companies for the financial support and encouragement they have provided to a variety of digital-literacy campaigns. These programs—ranging from classroom simulations for middle and high school students to videos by popular online personalities—are designed to educate users about questionable content, so they're less likely to be fooled by it or share it. Building on these admirable efforts, the companies should do more on this front.

Specifically, the platforms ought to make digital literacy lessons a permanent and prominently available feature on each of their sites. Doubtless, the companies would prefer not to remind users every time they log in that disinformation casts a shadow over social media. But that's the reality. The more often users are reminded of this fact—and are taught how to distinguish real from fake—the less influence false content will wield. Concise, incisive instruction, possibly presented in FAQs format, should be just one click away for all users of all of the platforms, all of the time.

Endnotes

- 1 Sarah Mervosh, "Distorted Videos of Nancy Pelosi Spread on Facebook and Twitter, Helped by Trump," *The New York Times*, May 24, 2019 (<https://www.nytimes.com/2019/05/24/us/politics/pelosi-doc-tored-video.html>). Another widely circulated "cheapfake" video falsely made it look as if CNN correspondent Jim Acosta was physically aggressive with a female White House intern during a November 2018 presidential press conference. We discussed the Acosta episode in our last report: Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do," New York University Stern Center for Business and Human Rights, March 2019 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_domestic_disinformation_digital?e=31640827/68184927).
- 2 Glenn Kessler, Salvador Rizzo, and Meg Kelly, "President Trump Has Made 12,019 False or Misleading Claims Over 928 Days," *The Washington Post*, August 12, 2019 (<https://www.washingtonpost.com/politics/2019/08/12/president-trump-has-made-false-or-misleading-claims-over-days/>).
- 3 Clint Watts, "The Kremlin's Strategy for the 2020 U.S. Election: Secure the Base, Split the Opposition," *Daily Beast*, May 29, 2019 (<https://www.thedailybeast.com/the-kremlins-strategy-for-the-2020-us-election-secure-the-base-split-the-opposition>).
- 4 James Loke Hale, "More than 500 Hours of Content Are Now Being Uploaded to YouTube Every Minute," *TubeFilter*, May 7, 2019 (<https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/>).
- 5 Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (supra note 1).
- 6 Paul M. Barrett, Tara Wadhwa, and Dorothée Baumann-Pauly, "Combating Russian Disinformation: The Case for Stepping Up the Fight Online," New York University Stern Center for Business and Human Rights, July 2018 (https://issuu.com/nyusterncenterforbusinessandhumanri/docs/nyu_domestic_disinformation_digital?e=31640827/68184927).
- 7 "Full Transcript: Robert Mueller's Statement on the Russia Investigation," *Politico*, May 29, 2019 (<https://www.politico.com/story/2019/05/29/robert-mueller-statement-russia-investigation-text-transcript-1346453>).
- 8 Dustin Volz, "No Significant Foreign Interference Seen on Midterm Vote," *The Wall Street Journal*, November 7, 2018 (<https://www.wsj.com/articles/u-s-on-alert-for-election-interference-says-nothing-significant-spotted-1541521048>).
- 9 Ellen Nakashima, "U.S. Cyber Command Operation Disrupted Internet Access of Russian Troll Farm on Day of 2018 Midterms," *The Washington Post*, February 27, 2019 (https://www.washingtonpost.com/world/national-security/us-cyber-command-operation-disrupted-internet-access-of-russian-troll-factory-on-day-of-2018-midterms/2019/02/26/1827fc9e-36d6-11e9-af5b-b51b7f322e9_story.html?utm_term=.85441f897537).
- 10 Interview with the author.
- 11 Interview with the author.
- 12 Eric Schmitt, David E. Sanger, Maggie Haberman, "In Push for 2020 Election Security, Top Official Was Warned: Don't Tell Trump," *The New York Times*, April 24, 2019 (<https://www.nytimes.com/2019/04/24/us/politics/russia-2020-election-trump.html>).
- 13 Daniel R. Coats, "Worldwide Threat Assessment of the U.S. Intelligence Community," Office of the Director of National Intelligence, January 29, 2019 (<https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf>).
- 14 Tom Van de Weghe, "Six Lessons from My Deepfakes Research at Stanford," *Medium*, May 29, 2019 (<https://medium.com/jsk-class-of-2019/six-lessons-from-my-deepfake-research-at-stanford-1666594a8e50>) and Leo Kelion, "Reddit Bans Deepfake Porn Videos," *BBC*, February 7, 2018 (<https://www.bbc.com/news/technology-42984127>).
- 15 John Villasenor, "Deepfakes, Social Media, and the 2020 Election," *TechTank*, Brookings Institution, June 3, 2019 (<https://www.brookings.edu/blog/techtank/2019/06/03/deepfakes-social-media-and-the-2020-election/>).
- 16 Samantha Putterman, "Zuckerberg Video about 'Billions of People's Stolen Data' Is a Deepfake," *Politifact*, June 12, 2019 (<https://www.politifact.com/facebook-fact-checks/statements/2019/jun/12/instagram-posts/zuckerberg-video-about-billions-peoples-stolen-dat/>).
- 17 Danielle Keats Citron, "Prepared Written Testimony and Statements for the Record," Hearing on the National Security Challenges of Artificial Intelligence, Manipulated Media, and "Deep Fakes," U.S. House Permanent Select Committee on Intelligence, June 13, 2019 (https://intelligence.house.gov/uploadedfiles/citron_testimony_for_house_committee_on_deep_fakes.pdf).
- 18 Ali Breland, "The Bizarre and Terrifying Case of the 'Deepfake' Video that Helped Bring an African Nation to the Brink," *Mother Jones*, March 15, 2019 (<https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>).
- 19 Robert Chesney, Danielle Keats Citron, Quinta Jurecic, "About That Pelosi Video: What to Do About 'Cheapfakes' in 2020," *Lawfare*, May 29, 2019 (<https://www.lawfareblog.com/about-pelosi-video-what-to-do-about-cheapfakes-2020>).
- 20 Interview with the author. It's worth noting that the mainstream media contribute to the dissemination of distorted videos.
- 21 Robert S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election," Department of Justice, March 2019 (<https://assets.documentcloud.org/documents/5955118/The-Mueller-Report.pdf>).
- 22 "Removing Bad Actors on Facebook," Facebook, July 31, 2018 (<https://newsroom.fb.com/news/2018/07/removing-bad-actors-on-facebook/>).
- 23 Joe Heim et al., "White-Supremacist Rally Near White House Dwarfed by Thousands of Anti-Hate Protesters," *The Washington Post*, August 12, 2018 (https://www.washingtonpost.com/local/washington-readies-for-todays-planned-white-supremacist-rally-near-white-house/2018/08/12/551720c4-9c28-11e8-8d5e-c6c594024954_story.html?utm_term=.569627792071).
- 24 Renee DiResta et al., "The Tactics and Tropes of the Internet Research Agency," New Knowledge for the U.S. Senate Select Committee on Intelligence, December 17, 2018 (<https://www.intelligence.senate.gov/press/new-reports-shed-light-internet-research-agency%E2%80%99s-social-media-tactics>).
- 25 Stan Schroeder, "U.S. Users Are Leaving Facebook, New Study Claims," *Mashable*, March 7, 2019 (<https://mashable.com/article/facebook-losing-users-us/>).

Endnotes (continued)

- 26 Alex Pasternack, "It's Not Over: Russia's Divisive Internet Memes Are Still Racking Up Likes," *Fast Company*, December 19, 2018 (<https://www.fastcompany.com/90283167/russia-instagram-war-face-book-memes>).
- 27 Taylor Lorenz, "Instagram is the Internet's New Home for Hate," *The Atlantic*, March 21, 2019 (<https://www.theatlantic.com/technology/archive/2019/03/instagram-is-the-internets-new-home-for-hate/585382/>).
- 28 Madhumita Murgia, Stephanie Findlay, and Andres Schipani, "India: The WhatsApp Election," *Financial Times*, May 5, 2019 (<https://www.ft.com/content/9fe88fba-6c0d-11e9-a9a5-351eeaf6d84>); Vinu Gopal and Sheera Frenkel, "In India Election, False Posts and Hate Speech Flummox Facebook," *The New York Times*, April 1, 2019 (<https://www.nytimes.com/2019/04/01/technology/india-elections-facebook.html>), and Philippa Williams and Lipika Kamra, "Technology Could Torpedo India's First WhatsApp Election," *Quartz*, March 4, 2019 (<https://qz.com/india/1563318/indias-2019-election-is-threatened-by-fake-news-on-whatsapp/>).
- 29 Interview with the author.
- 30 Alice Revelli and Lee Foster, "Network of Social Media Accounts Impersonates U.S. Political Candidates, Leverages U.S. and Israeli Media in Support of Iranian Interests," *FireEye*, May 28, 2019 (<https://www.fireeye.com/blog/threat-research/2019/05/social-media-network-impersonates-us-political-candidates-supports-iranian-interests.html>).
- 31 Nathaniel Gleicher, "Removing More Coordinated Inauthentic Behavior From Iran," Facebook, May 28, 2019 (<https://newsroom.fb.com/news/2019/05/removing-more-cib-from-iran/>).
- 32 Yoel Roth, Twitter, May 28, 2019 (<https://twitter.com/yoyoel/status/1133448387657736192>). All told, Twitter reportedly removed more than 7,000 phony accounts tied to Iran during the first seven months of 2019. See Craig Timberg and Tony Romm, "It's Not Just the Russians Anymore as Iranian and Others Turn Up Disinformation Efforts Ahead of 2020 Vote," *The Washington Post*, July 25, 2019 (https://www.washingtonpost.com/technology/2019/07/25/its-not-just-russians-anymore-iranians-others-turn-up-disinformation-efforts-ahead-vote/?utm_term=.4408c57407ff).
- 33 Sheera Frenkel and Nicholas Fandos, "Facebook Identifies New Influence Operations Spanning Globe," *The New York Times*, August 21, 2018 (<https://www.nytimes.com/2018/08/21/technology/facebook-political-influence-midterms.html>).
- 34 Interview with the author.
- 35 Interview with the author.
- 36 See, e.g., comments of Lisa-Maria Neudert: "Junk News Dominating Coverage of U.S. Midterms on Social Media, New Research Finds," Oxford Internet Institute, November 1, 2018 (<https://www.oii.ox.ac.uk/news/releases/junk-news-dominating-coverage-of-us-midterms-on-social-media-new-research-finds/>).
- 37 Cat Zakrzewski, "The Technology 202: Disinformation Spread by Americans Is 'the Hardest Challenge that We Have,' DHS Official Says," *The Washington Post*, April 12, 2019 (https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/04/12/the-technology-202-disinformation-spread-by-americans-is-the-hardest-challenge-that-we-have-dhs-official-says/5caf9cf91ad2e567949ec16c/?utm_term=.ecec91773e5f).
- 38 See, e.g., Vidya Narayanan et al., "Polarization, Partisanship, and Junk News Consumption Over Social Media in the U.S.," Oxford Internet Institute, February 6, 2018 (<https://comprop.oii.ox.ac.uk/research/polarization-partisanship-and-junk-news/>) and Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and the Radicalization of American Politics*, Oxford University Press, 2018 (<https://global.oup.com/academic/product/network-propaganda-9780190923631?cc=us&lang=en&>). See also, Paul M. Barrett, "Tackling Domestic Disinformation: What the Social Media Companies Need to Do" (supra note 1). As discussed in "Tackling Domestic Disinformation," a good deal of false content bubbles up from secondary social media sites, such as Reddit and Gab, to Twitter, YouTube, and Facebook.
- 39 "Candace Owens: Big Tech Trying to Ban Conservatives Ahead of 2020 Election," *Breitbart*, June 12, 2019 (<https://www.breitbart.com/politics/2019/06/12/watch-candace-owens-big-tech-trying-to-ban-conservatives-ahead-of-2020-election/>).
- 40 See, e.g., Kurt Wagner, "Facebook Bans Alex Jones, Milo Yiannopoulos, Other Far-Right Figures," *Bloomberg*, May 2, 2019 (<https://www.bloomberg.com/news/articles/2019-05-02/facebook-bans-alex-jones-yiannopoulos-other-far-right-figures>).
- 41 Lauren Feiner, "Trump Says Facebook, Amazon, and Google Were Colluding With Democrats Against Him," *CNN*, June 10, 2019 (<https://www.cnn.com/2019/06/10/trump-sounds-off-on-facebook-huawei-and-5g.html>).
- 42 Donald J. Trump, Twitter, July 11, 2019 (<https://twitter.com/realDonaldTrump/status/1149279675660869632>).
- 43 Ben Brody, "Trump Says He'll Summon Companies After Social Media 'Summit,'" *Bloomberg*, July 11, 2019 (<https://www.bloomberg.com/news/articles/2019-07-11/trump-s-social-media-summit-brings-fringe-voices-to-white-house>).
- 44 Caroline Kelly, "'It Won't Work': Kamala Harris' Campaign Slams Online Attacks on Her," *CNN*, June 30, 2019 (<https://www.cnn.com/2019/06/29/politics/kamala-harris-responds-donald-trump-jr-race-african-american-black/index.html>).
- 45 Adam K. Raymond, "Right-Wing Troll Jacob Wohl Allegedly Behind Fake Buttigieg Sexual-Assault Claim," *New York*, April 30, 2019 (<http://nymag.com/intelligencer/2019/04/right-wing-troll-set-up-fake-buttigieg-sexual-assault-claim.html>).
- 46 "A Progressive Activist Defends His Deceptive Tactics," *WNYC*, January 11, 2019 (<https://www.wnycstudios.org/story/progressive-activist-defends-his-deceptive-tactics>).
- 47 Scott Shane and Alan Blinder, "Democrats Faked Online Push to Outlaw Alcohol in Alabama Race," *The New York Times*, January 7, 2019 (<https://www.nytimes.com/2019/01/07/us/politics/alabama-senate-facebook-roy-moore.html>).
- 48 Tony Romm, Elizabeth Dwoskin, and Craig Timberg, "Facebook Is Investigating the Political Pages and Ads of Another Group Backed by Reid Hoffman," *The Washington Post*, January 7, 2019 (https://www.washingtonpost.com/technology/2019/01/07/facebook-is-investigating-political-pages-ads-another-group-backed-by-reid-hoffman/?utm_term=.e1f7350447ac).

- 49 “A Progressive Activist Defends His Deceptive Tactics” (supra note 46).
- 50 Interview with the author.
- 51 Tony Romm and Craig Timberg, “Facebook Suspends Five Accounts, Including That of a Social Media Researcher, for Misleading Tactics in Alabama Elections,” *The Washington Post*, December 22, 2018 (https://www.washingtonpost.com/technology/2018/12/22/facebook-suspends-five-accounts-including-social-media-researcher-misleading-tactics-alabama-election/?utm_term=.6fd96ea070dd) and Scott Shane and Alan Blinder, “Democrats Faked Online Push to Outlaw Alcohol in Alabama Race” (supra note 47).
- 52 Andy Greenberg, “Alphabet-Owned Jigsaw Bought a Russian Troll Campaign as an Experiment,” *Wired*, June 12, 2019 (<https://www.wired.com/story/jigsaw-russia-disinformation-social-media-stalin-alphabet/>).
- 53 Adam Entous and Ronan Farrow, “Private Mossad for Hire,” *The New Yorker*, February 11, 2019 (<https://www.newyorker.com/magazine/2019/02/18/private-mossad-for-hire>) and Mark Mazzetti, et al., “Rick Gates Sought Online Manipulation Plans from Israeli Intelligence Firm for Trump Campaign,” *The New York Times*, October 8, 2018 (<https://www.nytimes.com/2018/10/08/us/politics/rick-gates-psy-group-trump.html>).
- 54 Nathaniel Gleicher, “Removing Coordinated Inauthentic Behavior from Israel,” Facebook, May 16, 2019 (<https://newsroom.fb.com/news/2019/05/removing-coordinated-inauthentic-behavior-from-israel/>).
- 55 Isabel Debre and Raphael Satter, “Facebook Busts Israel-based Campaign to Disrupt Elections,” *Associated Press*, May 16, 2019 (<https://www.apnews.com/7d334cb8793f49889be1bbf89f47ae5c>) and Simona Wineglass, “Who Is Behind Israel’s Archimedes Group, Banned by Facebook for Election Fakery?” *The Times of Israel*, May 19, 2019 (<https://www.timesofisrael.com/who-is-behind-israels-archimedes-group-banned-by-facebook-for-election-fakery/>).
- 56 Robert S. Mueller III, “Report on the Investigation into Russian Interference in the 2016 Presidential Election” (supra note 21).
- 57 “Facebook Press Call,” Facebook, April 26, 2019 (https://fbnewsroom.us.files.wordpress.com/2019/04/transcript-9549757_final.pdf).
- 58 Young Mie Kim, “Voter Suppression has Gone Digital,” New York University Brennan Center for Justice, November 20, 2018 (<https://www.brennancenter.org/blog/voter-suppression-has-gone-digital>).
- 59 Christopher Bing, “Exclusive: Twitter Deletes Over 10,000 Accounts that Sought to Discourage U.S. Voting,” *Reuters*, November 2, 2018 (https://www.reuters.com/article/us-usa-election-twitter-exclusive/exclusive-twitter-deletes-over-10000-accounts-seeking-to-discourage-voting-idUSKCN1N72FA?feedType=RSS&feedName=technologyNews&utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+reuters%2FtechnologyNews+%28Reuters+Technology+News%29).
- 60 Guy Rosen, “An Update on How We Are Doing at Enforcing Our Community Standards,” Facebook, May 23, 2019 (<https://newsroom.fb.com/news/2019/05/enforcing-our-community-standards-3/>).
- 61 Nathaniel Gleicher, “Coordinated Inauthentic Behavior Explained,” Facebook, December 6, 2018 (<https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>).
- 62 Interview with the author.
- 63 Henry Silverman, “The Next Phase in Fighting Misinformation,” Facebook, April 10, 2019 (<https://newsroom.fb.com/news/2019/04/tackling-more-false-news-more-quickly/>).
- 64 “Zuckerberg Says It’s ‘Really Important’ for Facebook to Form Deepfake Video Policy,” *CBS News*, June 27, 2019 (<https://www.cbsnews.com/news/facebook-zuckerberg-says-its-really-important-to-form-policy-on-deepfake-videos/>).
- 65 Arjun Kharpal, “Facebook Doesn’t Want to Be the ‘Arbiter of the Truth,’ Top Executive Sheryl Sandberg Says Amid Fake News Criticism,” *CNBC*, April 24, 2017 (<https://www.cnn.com/2017/04/24/facebook-fake-news-sheryl-sanberg.html>).
- 66 “Cooper Grills Facebook VP for Keeping Pelosi Video Up,” *CNN*, May 25, 2019 (<https://www.cnn.com/videos/tech/2019/05/25/facebook-monika-bickert-pelosi-video-cooper-intv-sot-ac360-vpx.cnn>).
- 67 Jessica Leinwand, “Expanding Our Policies on Voter Suppression,” Facebook, October 15, 2018 (<https://newsroom.fb.com/news/2018/10/voter-suppression-policies/>) and Sheryl Sandberg, “A Second Update on Our Civil Rights Audit,” Facebook, June 30, 2019 (<https://newsroom.fb.com/news/2019/06/second-update-civil-rights-audit/>).
- 68 Kevin Kane, “Securing U.S. Election Infrastructure and Protecting Political Discourse,” U.S. House Committee on Oversight and Reform, May 22, 2019 (<https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-KaneK-20190522.pdf>) and “Policies and Safety,” YouTube, undated (https://support.google.com/youtube/answer/2801973?hl=en&ref_topic=9282365).
- 69 “Our Ongoing Work to Tackle Hate,” YouTube, June 5, 2019 (<https://youtube.googleblog.com/2019/06/our-ongoing-work-to-tackle-hate.html>).
- 70 “Continuing Our Work to Improve Recommendations on YouTube,” YouTube, January 25, 2019 (<https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html>).
- 71 Yoel Roth, “Information Operations on Twitter: Principles, Process, and Disclosure,” Twitter, June 13, 2019 (https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html).
- 72 Henry Silverman, “The Next Phase in Fighting Misinformation” (supra note 63).
- 73 Sam Levin and Julia Carrie Wong, “‘He’s Learned Nothing’: Zuckerberg Floats Crowdsourcing Facebook Fact-Checks,” *The Guardian*, February 20, 2019 (<https://www.theguardian.com/technology/2019/feb/20/facebook-fact-checking-crowdsourced-mark-zuckerberg>). Snopes said it ended its partnership with Facebook because the arrangement was too time-consuming and inefficient. See, e.g., Daniel Funke, “Snopes Pulls Out of its Fact-Checking Partnership with Facebook,” *Poynter*, February 1, 2019 (<https://www.poynter.org/fact-checking/2019/snopes-pulls-out-of-its-fact-checking-partnership-with-facebook/>).
- 74 “Interview with Facebook CEO Mark Zuckerberg,” *ABC News*, April 4, 2019 (<https://abcnews.go.com/Business/interview-facebook-ceo-mark-zuckerberg-transcript/story?id=62152829>).
- 75 Kevin Kane, “Securing U.S. Election Infrastructure and Protecting Political Discourse” (supra note 68).

Endnotes (continued)

- 76 Richard Salgado, “Securing U.S. Election Infrastructure and Protecting Political Discourse,” U.S. House Committee on Oversight and Reform, May 22, 2019 (<https://docs.house.gov/meetings/GO/GO06/20190522/109538/HHRG-116-GO06-Wstate-SalgadoR-20190522.pdf>).
- 77 Katie Harbath and Samidh Chakrabarti, “Expanding Our Efforts to Protect Elections in 2019,” Facebook, January 28, 2019 (<https://newsroom.fb.com/news/2019/01/elections-2019/>).
- 78 Samidh Chakrabarti, “Fighting Election Interference in Real Time,” Facebook, October 18, 2018 (<https://newsroom.fb.com/news/2018/10/war-room/>).
- 79 Sheera Frenkel and Matthew Rosenberg, “Top Tech Companies Met with Intelligence Officials to Discuss Midterms,” *The New York Times*, June 25, 2018 (<https://www.nytimes.com/2018/06/25/technology/tech-meeting-midterm-elections.html>) and “Interview with Facebook CEO Mark Zuckerberg” (supra note 74).
- 80 Kevin Kane, “Securing U.S. Election Infrastructure and Protecting Political Discourse” (supra note 68).
- 81 Del Harvey and Yoel Roth, “An Update on Our Elections Integrity Work,” Twitter, October 1, 2018 (https://blog.twitter.com/en_us/topics/company/2018/an-update-on-our-elections-integrity-work.html).
- 82 See, e.g., Del Harvey and Bruce Falck, “Announcing New U.S. Issue Ads Policy,” Twitter, August 30, 2018 (https://blog.twitter.com/en_us/topics/company/2018/Announcing-new-US-issue-ads-policy.html). Researchers have reported technical problems with Facebook’s advertising database: See, Matthew Rosenberg, “Ad Tool Facebook Built to Fight Disinformation Doesn’t Work as Advertised,” *The New York Times*, July 25, 2019 (<https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html>).
- 83 “Announcement: RT and Sputnik Advertising,” Twitter, October 26, 2017 (https://blog.twitter.com/en_us/topics/company/2017/Announcement-RT-and-Sputnik-Advertising.html).
- 84 William Turton, “We Posed as 100 Senators to Run Ads on Facebook. Facebook Approved Them All,” *Vice News*, October 30, 2018 (https://news.vice.com/en_us/article/xw9n3q/we-posed-as-100-senators-to-run-ads-on-facebook-facebook-approved-all-of-them).
- 85 Jordain Carney, “Democrats Detail New Strategy to Pressure McConnell on Election Security Bills,” *The Hill*, June 18, 2019 (<https://thehill.com/blogs/floor-action/senate/449174-democrats-detail-new-strategy-to-pressure-mcconnell-on-election>) and Maggie Miller, “Bipartisan Group of Senators Seeks to Increase Transparency of Online Political Ads,” *The Hill*, May 8, 2019 (<https://thehill.com/homenews/senate/442598-bipartisan-group-of-senators-seek-to-increase-transparency-of-online>).
- 86 Nahema Marchal, et al., “Polarization, Partisanship, and Junk News Consumption on Social Media During the 2018 U.S. Midterm Elections,” Oxford Internet Institute, November 1, 2018 (http://blogs.oxi.ox.ac.uk/comprop/wp-content/uploads/sites/93/2018/11/marchal_et_al.pdf).
- 87 Andrew Guess, et al., “Fake News, Facebook Ads, and Misperceptions,” University of Michigan, February 2019 (<http://www-personal.umich.edu/~bnyhan/fake-news-2018.pdf>).
- 88 Hunt Allcott, Matthew Gentzkow, and Chuan Yu, “Trends in the Diffusion of Misinformation on Social Media,” Stanford University, October 2018 (<https://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf>).
- 89 “Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee, and the Committee of the Regions,” European Commission, June 14, 2019 (https://eeas.europa.eu/sites/eeas/files/joint_report_on_disinformation.pdf).
- 90 Nahema Marchal, et al., “Junk News During the EU Parliamentary Elections: Lessons from a Seven-Language Study of Twitter and Facebook,” Oxford Internet Institute, May 21, 2019 (<https://comp.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Data-Memo.pdf>).
- 91 “New Research Shows Facebook Making Strides Against False News,” Facebook, February 7, 2019 (<https://newsroom.fb.com/news/2018/10/inside-feed-michigan-lemonade/>).
- 92 “Truth on the Ballot: Fraudulent News, the Midterm Elections, and Prospects for 2020,” PEN America, March 13, 2019 (<https://pen.org/wp-content/uploads/2019/03/Truth-on-the-Ballot-report.pdf>).
- 93 Daniel Benjamin and Steven Simon, “How Fake News Could Lead to a Real War,” *Politico Magazine*, July 5, 2019 (<https://www.politico.com/magazine/story/2019/07/05/fake-news-real-war-227272>).
- 94 Robert Chesney and Danielle Keats Citron, “Deepfakes: A Looming Challenge for Privacy, Democracy, and National Security,” *California Law Review*, forthcoming, 2019 (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954).
- 95 Daisy Stanton, “Advancing Research on Fake Audio Detection,” Google, January 31, 2019 (<https://www.blog.google/outreach-initiatives/google-news-initiative/advancing-research-fake-audio-detection/>).
- 96 “Community Guidelines,” Pinterest, undated (<https://policy.pinterest.com/en/community-guidelines>).
- 97 Casey Newton, “Instagram Is Facing a Reckoning Over Misinformation,” *The Verge*, March 22, 2019 (<https://www.theverge.com/instagram/2019/3/22/18276479/instagram-reckoning-anti-vaxx-misinformation-conspiracy>).
- 98 Paris Martineau, “Instagram Can Find Misleading Posts—but Won’t Take Them Down,” *Wired*, May 8, 2019 (<https://www.wired.com/story/instagram-find-misleading-posts-wont-take-down/>).
- 99 “More Changes to Forwarding,” WhatsApp, January 21, 2019 (<https://blog.whatsapp.com/10000647/More-changes-to-forwarding>).
- 100 Luiza Bandeira, et al., “Inauthentic Israeli Facebook Assets Target the World,” Digital Forensic Research Lab, May 17, 2019 (<https://medium.com/dfrlab/inauthentic-israeli-facebook-assets-target-the-world-281ad7254264>).
- 101 Jessica Brandt and Bradley Hanlon, “Online Information Operations Cross Platforms. Tech Companies’ Responses Should, Too,” *Lawfare*, April 26, 2019 (<https://www.lawfareblog.com/online-information-operations-cross-platforms-tech-companies-responses-should-too>).

NYU Stern Center for Business and Human Rights
Leonard N. Stern School of Business
44 West 4th Street, Suite 800
New York, NY 10012
+1 212-998-0261
bhr@stern.nyu.edu
bhr.stern.nyu.edu

© 2019 NYU Stern Center for Business and Human Rights
All rights reserved. This work is licensed under the
Creative Commons Attribution-NonCommercial 4.0
International License. To view a copy of the license,
visit <http://creativecommons.org/licenses/by-nc/4.0/>.



Center for Business
and Human Rights



CÂMARA DOS DEPUTADOS
Gabinete da Deputada **Lídice da Mata** - PSB/BA

Ofício GDLMAT nº 074/2019

Brasília, 23 de setembro de 2019.

Exmo. Senhor Presidente,

Com meus cumprimentos, requeiro à Vossa Excelência que seja solicitado à Secretaria da Comissão Parlamentar Mista de Inquérito (CPMI) das Fake News providências para imediata e urgente tradução para o Português de dois documentos anexados, que tratam de temas de interesse e que visam instruir os trabalhos dos membros da referida Comissão.

Os documentos referidos são: 1) **“Disinformation and ‘fake news’: Final Report”** - Desinformação e 'notícias falsas': relatório final, produzido pela House of Commons Digital, Culture, Media and Sport Committee (Câmara dos Comuns Comitê Digital, Cultura, Mídia e Esporte), em fevereiro de 2019; e 2) **“Disinformation and the 2020 Election: How the Social Media Industry Should Prepare”** - “Desinformação e eleição de 2020: como a indústria das mídias sociais deve se preparar”, de autoria de Paul M. Barret, publicado por Center for Business and Human Rights (Centro de Negócios e Direitos Humanos) em setembro de 2019.

Dada à relevância desses estudos para os trabalhos da relatoria da referida Comissão, solicitamos urgência na tradução.

Respeitosamente,


Lídice da Mata

Deputada Federal PSB/BA
Relatora da CPMI das Fake News

À Sua Excelência o Senhor
SENADOR ANGELO CORONEL
Presidente da CPMI – Fake News

Recebido em
25/9/19
