‘Deepfakes’ atingem principalmente mulheres, alerta especialista

De acordo com Sam Gregory, debate sobre vídeos adulterados por inteligência artificial deve focar em público mais vulnerável a ameaças de manipulação

(Estadão, 20/10/2019 – acesse no site de origem)

Antes do pleito de 2018, havia o temor que os chamados deepfakes, vídeos produzidos com inteligência artificial para simular a aparência de uma pessoa, interferissem no processo eleitoral. No entanto, as eleições passaram e esse tipo de conteúdo falso continuou restrito à pornografia, em que os rostos de celebridades e outras mulheres são inseridos em cenas de sexo.

O especialista em tecnologia Sam Gregory, diretor da organização Witness, alerta que é necessário centrar a discussão sobre deepfakes em proteger as vítimas mais afetadas hoje: as mulheres. Ele também defende que o debate em torno do assunto não fique restrito à Europa e aos Estados Unidos, e que países do hemisfério sul, como o Brasil, também proponham soluções

O especialista em tecnologia Sam Gregory em seminário promovido pela Associação Nacional de Jornais (ANJ). Foto: Witness

Gregory participou do seminário “Desinformação: Antídotos e Tendências”, promovido pela Associação Nacional de Jornais (ANJ), na quinta-feira, 17.

Você alertou que as mulheres estão mais vulneráveis às deepfakes. Por que isso acontece?

Uma pesquisa recente do grupo DeepTrace Labs mostrou que cerca de 96% dos deepfakes são pornográficos ou imagens de sexo não-consensual. Muitos são de celebridades e pessoas públicas, mas há também de indivíduos privados. Não deve ser surpresa que essa tecnologia se tornou uma arma contra mulheres, porque elas são sujeitas à maior parte do assédio online, de qualquer tipo. Devemos pensar em soluções que enfrentem a violência de gênero, e que usem uma perspectiva legal e das plataformas. E não podemos ser complacentes com esse problema. Por exemplo, lançaram um aplicativo para fazer mulheres parecerem estar nuas. Devemos denunciar isso como um uso inaceitável da tecnologia. Há muitas pessoas que minimizam como “apenas” um problema de violência de gênero, pois ainda não tem sido usado na política. Esse é um pensamento horrível. Observando as ameaças existentes, é fácil prever como isso pode se expandir para ataques mais amplos, contra jornalistas e ativistas, e contra a integridade das evidências em vídeo.

Você apontou que, assim como é com as mulheres, a ameaça pode evoluir para populações mais vulneráveis.

O pilar do nosso trabalho com deepfakes é aprender com a forma com que enfrentamos outros

problemas da desinformação. As soluções não focam nas pessoas que realmente foram atingidas: mulheres, populações vulneráveis, comunicadores comunitários, jornalistas que trabalham sem estrutura de apoio. É muito importante nos concentrarmos nas pessoas que são mais prejudicadas. Devemos nos certificar de que não estamos só ouvindo Washington, Bruxelas e o Vale do Silício. O problema não está centrado na política dos Estados Unidos. Isso é entender errado o escopo da desinformação hoje.

Políticos locais também estão menos protegidos do que candidatos com visibilidade nacional. Devemos nos preocupar com as deepfakes nas eleições 2020?

Não devemos fazer previsões de que as deepfakes serão um problema em 2020. Temos que ter cuidado para não causar pânico. Não queremos que as pessoas pensem que haverá deepfakes em todos os lugares em 2020, porque provavelmente não haverá. Teremos desinformação, mas é improvável que tenhamos deepfakes em larga escala. O que devemos fazer é nos preparar melhor, caso isso aconteça. Essa preparação valerá a pena quando as deepfakes se tornarem mais difundidas. Como a tecnologia está melhorando, se tornando mais acessível e mais fácil de usar, é muito provável que no futuro seja usada de forma maliciosa em larga escala. Para as eleições 2020 no Brasil e nos Estados Unidos, devemos focar em preparar jornalistas e checadores com as ferramentas que eles precisam. Eles estarão na linha de frente para fazer checagens, e no momento não há muitos recursos para isso.

O que os políticos podem fazer em relação às deepfakes?

É importante que os políticos se comprometam a não usar essa ferramenta. Deveria ser uma norma não compartilhar vídeos manipulados em campanha. Vimos isso acontecendo na Europa: muitos candidatos ao parlamento europeu firmaram esse compromisso. Mas não vimos nos Estados Unidos, e nem sabemos se há tentativas nesse sentido no Brasil.

As plataformas vão firmar um compromisso de colaborar com o programa de combate à desinformação do Tribunal Superior Eleitoral (TSE). As empresas de tecnologia estão entendendo melhor sua responsabilidade nesse contexto?

Espero que elas estejam. No contexto das deepfakes, as plataformas parecem estar se preparando melhor. Estão construindo ferramentas de detecção melhores, e disponibilizando para mais pessoas, pensando em como podem colaborar com diferentes grupos como a mídia ou o governo. Mas, na esfera mais ampla, há pedidos de acadêmicos, pesquisadores e entidades do governo por mais transparência. Em geral, as plataformas não oferecem transparência o suficiente para permitir que os jornalistas decidam de forma correta o que investigar online. E não incluíram ferramentas de checagem no WhatsApp para ajudar os usuários a entender o que é falso. As plataformas definitivamente podem avançar mais.

Captura de tela do FakeApp com fotos do ator Jake Gyllenhaal e de Kevin Roose, repórter do The New York Times Foto: Handout via The New York Times

Em sua palestra, você mencionou que é possível treinar um algoritmo a manipular expressões faciais com apenas 16 selfies. Certamente já publicamos muito mais que isso nas redes sociais. O cidadão comum pode se proteger de abusos?

É muito difícil para uma pessoa comum evitar ter fotos online. Especialmente se você

considera um contexto maior, com todas as câmeras de vigilância que existem. Pela forma com que a capacidade de criar mídia sintética e deepfakes evolui, a tendência é requerer menos dados de treinamento (imagens para que o algoritmo aprenda a imitar uma pessoa), e ter modelos que não precisem ser treinados com imagens de uma pessoa específica. Há discussões centradas em dar de volta autonomia às pessoas, para que elas possam controlar suas próprias imagens. Há uma proposta interessante de usar adversarial perturbations (perturbações contraditórias). É um tipo de mudança invisível em uma imagem que faz com que seja mais difícil que o computador reconheça que é você. Basicamente, é uma forma de enganar um sistema de inteligência.

Por exemplo, você poderia inserir essas "perturbações" em todas as suas imagens para impedir que elas sejam buscadas e usadas para treinar um algoritmo de deepfake. Como defensor de direitos humanos, gosto de analisar essas opções, que permitem a opção de ser mais anônimo, mais invisível. Mas ainda há desafios técnicos.

De qualquer forma, é importante que as pessoas entendam que suas imagens podem ser usadas dessa forma.

As pessoas devem saber que suas imagens online são usadas para alimentar fakes. Mas, como defensor de direitos humanos e uma pessoa que pensa muito sobre segurança digital, acho que há muitas implicações de ter imagens online que são igualmente importantes — por exemplo, possibilitar que você seja identificado e rastreado no mundo real. É importante fazer com que as plataformas deem opções para as pessoas que querem optar por não ter suas imagens online. No futuro, poderíamos ter essas perturbações adversariais no Facebook? Podemos ter ferramentas para ajudar as pessoas a se tornarem anônimas?

Por Alessandra Monnerat