AS VERDADES E MENTIRAS DE UM *DEEPFAKE*

Os *deepfakes* dizem respeito à Inteligência Artificial (IA) utilizada para criar perfis, vídeos, imagens e áudios falsos de pessoas reais. Inicialmente, limitados a especialistas em Tecnologia, as ferramentas de *deepfake* são agora acessíveis globalmente a toda a população, a um custo irrisório e de fácil manuseamento. Além desta facilidade inerente à sua utilização e custo, as ferramentas que permitem a criação de conteúdo sintético são também cada vez mais.

Nas redes sociais começa a fazer-se sentir o impacto desta proliferação dos *deepfakes*. Com certeza já aconteceu à grande maioria de nós depararmo-nos com uma imagem ou vídeo *online*, que nos pareceu real, mas no qual ao mesmo tempo notamos algo de estranho, ou por um detalhe ou pelo próprio contexto. Também certamente já nos deparámos, nesses casos, com comentários de utilizadores a questionar se a imagem ou vídeo em questão, seria real ou produzida por IA. Esta atmosfera de dúvida parece-nos ser já uma pequena (mas grande) amostra daquilo que será um futuro próximo, marcado pela dificuldade de validar a veracidade daquilo que nos chega, daquilo que os nossos olhos veem, que os nossos ouvidos ouvem. No fundo, começamos a ter de questionar os nossos próprios sentidos. É fácil percebermos o impacto deste fenómeno em termos de disseminação de conteúdo falso, ainda mais na esfera cibernética, em que o conteúdo viaja e se propaga à velocidade da luz.

Os últimos anos têm sido marcados por avanços significativos na qualidade e realismo dos *deepfakes* cujos algoritmos têm vindo a aprimorar a capacidade de representação das leis físicas, desde pequenos detalhes de movimento, a sombras. Se, atualmente, são já difíceis de distinguir, com os avanços do futuro custa até imaginar o grau de dificuldade de distinção que se aproxima. Essa proliferação de conteúdo *deepfake* aliado ao desafio da deteção levantará questões sobre a autenticidade de qualquer conteúdo que encontraremos *online* e terá ainda o potencial de minar a confiança na informação tal como a compreendemos nos dias de hoje.

Mas os *deepfakes* não são só utilizados para espalhar informações falsas *online* e manipular a população, são também cada vez mais aplicados em ataques de engenharia social, tornando-os muito mais difíceis de detetar sendo por isso uma das tendências mais perigosas do cibercrime. Os cibercriminosos já começaram a transformar *deepfakes*

«Esta atmosfera de dúvida parece-nos ser já uma pequena (mas grande) amostra daquilo que será um futuro próximo, marcado pela dificuldade de validar a veracidade daquilo que nos chega, daquilo que os nossos olhos veem, que os nossos ouvidos ouvem. No fundo, começamos a ter de questionar os nossos próprios sentidos.»

«Os cibercriminosos já começaram a transformar *deepfakes* em armas de ciberataque a indivíduos, por exemplo ao enviarem mensagens de voz que parecem ser de uma pessoa conhecida, ou provindas de organizações, ao ludibriar sistemas biométricos e de reconhecimento facial, ou ainda ao imitar personalidades reais (como um executivo, um cliente, ou um parceiro da organização).»

em armas de ciberataque a indivíduos, por exemplo ao enviarem mensagens de voz que parecem ser de uma pessoa conhecida, ou provindas de organizações, ao ludibriar sistemas biométricos e de reconhecimento facial, ou ainda ao imitar personalidades reais (como um executivo, um cliente, ou um parceiro da organização). As intenções podem ser várias, desde ganhos financeiros a roubo de dados. À medida que a inovação veio dar um novo impulso ao cibercrime, através de técnicas de exploração de ameaças cibernéticas altamente inovadoras, podemos assumir que agora, mais do que nunca, o comportamento humano persistirá como uma vulnerabilidade que pode ser explorada pelo cibercrime, de forma ainda mais eficaz e disruptiva. Efetivamente, a ameaça dos *deepfakes* não provém somente da tecnologia utilizada para os criar, mas da tendência natural das pessoas para acreditarem no que veem e, como resultado, os *deepfakes* não precisam de ser particularmente avançados ou credíveis para serem eficazes na disseminação de informações falsas e/ou fraudulentas.

As empresas, ao lançarem estas ferramentas, deverão ter o cuidado de incorporar salvaguardas como marcas d'água digitais – ainda que isso provavelmente não seja suficiente. A realidade é que, de uma forma geral, a tecnologia utilizada para identificação e deteção de conteúdo *deepfake* ainda não é capaz de fornecer a maturidade e eficácia suficiente, e os atuais detetores enfrentam desafios enormes, principalmente, devido a dados incompletos ou dispersos. Para enfrentar este desafio, as organizações podem aplicar uma variedade de técnicas, como por exemplo, a análise forense de vídeos, a verificação de autenticidade baseada em *blockchain*, e o recurso a redes neurais (método de IA que ensina computadores a processar dados de uma forma inspirada pelo cérebro humano) especializadas em deteção de manipulações de conteúdos de Media.

Também ao nível governamental os líderes têm um papel importante, nomeadamente, ao criminalizar a utilização dos *deepfakes* usados para fins de roubo de dados, fraudes, desinformação, entre outros. Por exemplo, recentemente, em abril deste ano, o Governo do Reino Unido anunciou que a criação de imagens *deepfake* sexualmente explícitas será considerada crime em Inglaterra e no País de Gales ao abrigo de uma nova lei. Esperamos que este seja apenas o início de uma série de iniciativas a tomar pelos Estados, por todo o Mundo, e que incluam a criminalização de todas as formas de utilização maliciosa de *deepfake*.

No futuro, enfrentaremos uma paisagem digital onde os atores maliciosos já conseguem criar imagens, vídeos e até mesmo áudios falsos de maneira convincente. Este fenómeno, aliado à globalização do ciberespaço e à tendência natural humana de crença nos seus sentidos, levanta sérios desafios para os indivíduos, organizações e Estados. Além do trabalho das organizações e dos Estados que deve ser levado a cabo para combater este desafio, os indivíduos, enquanto partes integrantes da sociedade civil devem investir acima de tudo na sua literacia digital. Conhecimento é efetivamente poder, principalmente no que diz respeito a saber distinguir as verdades e a mentiras por detrás dos *deepfakes*.

Qual a anatomia de um *deepfake*?

Recolha e Preparação de Dados

Nesta etapa, o agressor recolhe os dados sobre o indivíduo-alvo, incluindo imagens, vídeos e gravações de áudio. Quantos mais dados forem recolhidos, mais preciso e certeiro o *deepfake* poderá ser.

Treino, manipulação e síntese

Utilização de *Deep Learning* para aprender os padrões, características e especificidades do indivíduo-alvo. Os modelos geram imagens sintéticas, vídeos ou áudios, que imitam a aparência e o comportamento do alvo. Os algoritmos ajustam expressões faciais, movimentos labiais, tom de voz, gestos e outras características do alvo. O conteúdo manipula-

«No futuro, enfrentaremos uma paisagem digital onde os atores maliciosos já conseguem criar imagens, vídeos e até mesmo áudios falsos de maneira convincente. Este fenómeno, aliado à globalização do ciberespaço e à tendência natural humana de crença nos seus sentidos, levanta sérios desafios para os indivíduos, organizações e Estados.»

do é posteriormente sintetizado.

Distribuição e Impacto

Depois de o *deepfake* ser criado, é distribuído por vários canais, como social media, websites ou plataformas de mensagens. *Deepfakes* podem ser usados para vários fins, incluindo espalhar informações erradas, criar vídeos falsos de personalidades públicas ou corporativas e, ainda, cometer fraudes. O impacto dos *deepfakes* pode variar, desde gerar a confusão/ caos público, a danos reputacionais e perdas financeiras avultadas, passando também pelo roubo de dados confidenciais.

Como prevenir os *deepfakes*?

A dramática proliferação de *deepfakes* começou a minar a nossa capacidade de discernimento da realidade ao obrigar a população a questionar algo que era tão certo como a informação apercebida através dos nossos sentidos.

Apesar de não existir uma solução imediata, é fundamental que todos os membros de uma organização, sem exceção, estejam conscientes do risco e o saibam reconhecer.

Bruno Castro

Fundador & CEO da VisionWare, Especialista em Cibersegurança e Investigação Forense

A EQUIPA DA VISIONWARE SUGERE 7 DICAS E CONSELHOS ÚTEIS DE COMO PODERÁ DETETAR *DEEPFAKES*:

1. Incongruências na pele e em partes do corpo

2. Sombras à volta dos olhos

3. Padrões de pestanejo invulgares

4. Brilho invulgar nos óculos

5. Movimentos labiais incompatíveis ou irrealistas

6. Coloração não natural dos lábios em relação ao rosto

7. Manchas irrealistas no rosto

O QUE AS ORGANIZAÇÕES PODEM FAZER PARA MINIMIZAR RISCOS?

• Desenvolver uma cultura de cibersegurança e ciber higiene dentro da organização (através de formações e treino contínuos).

• Autenticação forte e fortalecer a confirmação de identidade (biometria, autenticação em dois fatores, *passwords* fortes, modelo *zero-trust*).

• Strategic Intelligence e Análise de Risco (identificar indicadores-chave de risco, quantificar o impacto financeiro dos riscos de segurança.

• Investimento em Cibersegurança by Design.

• Investir em tecnologias avançadas para melhor monitorização, deteção e mitigação.

Nota: Alguns dados foram obtidos via monitorização da equipa Strategic Intelligence & Risk Analysis da VisionWare.