

El auge de las *deep fakes*: del entretenimiento a la desinformación

Ignacio Cascarrón y Sandra Vilches explican en *Newtral*¹ cómo se elaboran las denominadas *deepfakes* audiovisuales, cuya utilización puede ir desde el humor y el entretenimiento o la publicidad, hasta la desinformación en internet, incrementando ese riesgo sistémico.

Los autores, citando a algunos expertos, señalan cómo con algunas aplicaciones básicas de móvil y unas pocas fotografías pueden elaborarse esas *deepfakes* de un modo bastante eficaz, aunque sin llegar a los niveles de simulación que ofrece contar con un equipo especializado y técnicas avanzadas.

Entre estas técnicas se señalan:

- Las redes de difusión, que parten de una determinada imagen añadiéndole ruido (distorsionándola) hasta hacerla irreconocible. Después, el modelo realiza el proceso contrario, a partir de la imagen distorsionada, deshaciendo el ruido generado y volviendo a recrear una nueva imagen que no coincide con la original, pero tiene datos similares.
- Los modelos de consistencia (*consistency models*), que mejoran los resultados logrados por las redes de difusión, diseñados para convertir directamente el ruido en datos, y que necesitan menos entrenamiento que dichas redes

Las redes de difusión han ido sustituyendo a sus predecesoras, las redes generativas antagónicas (o GAN, *generative adversarial network*), que eran el modelo más novedoso hace sólo dos años con métodos como:

El **lip-syncing**, uno de los más sencillos y frecuentes para crear *deepfakes* audiovisuales, consistente en la sincronización de labios. Combina la clonación de voz con una técnica de edición de vídeos, haciendo que el movimiento de la boca de la persona suplantada coincida con el discurso verbal elegido o creado al efecto.

El **Face-swapping**, permite construir una cara a partir de rasgos de caras originales, a partir de la superposición de ambas imágenes. Cuantos más datos tenga el algoritmo y más variados sean (diferentes ángulos, expresiones faciales y grados de iluminación), más creíble será el *deepfake*. Puede combinarse con el uso del audio sintético, algo muy popular en las redes sociales.

¹ https://www.newtral.es/deepfakes-audiovisuales/20241022/?utm_medium=Social&utm_campaign=Echobox-Twitter&utm_source=Twitter#Echobox=1733409743