THREEPOINTS
THE SCHOOL FOR DIGITAL BUSINESS

Deepfake, ¿el fin de la realidad?

Carlos Rodríguez Abellán
Senior Data Scientist en Telefónica y
profesor colaborador de Three Points

Marzo, 2021

Partner Académico:

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Una escuela de:

OBS Business School

Autor

Carlos Rodríguez Abellán

- Científico de datos en Telefónica
- Experto en proyectos de analítica avanzada e inteligencia artificial
- Ingeniero en Tecnologías y Servicios de Telecomunicación por la Universidad Politécnica de Madrid
- MSc in Signal Processing and Machine *Learning* por la Universidad Politécnica de Madrid
- Profesor de Three Points


The
Elizabet Warren's plan
to rem ke American
Next!
Polluters;
HEALTH
INNOVATION
ISSUE
WHO GETS
TO BE
HEALTHY
WHAT WOMEN NEED
THE
Documents
Par

Índice

Capítulo 1

Introducción

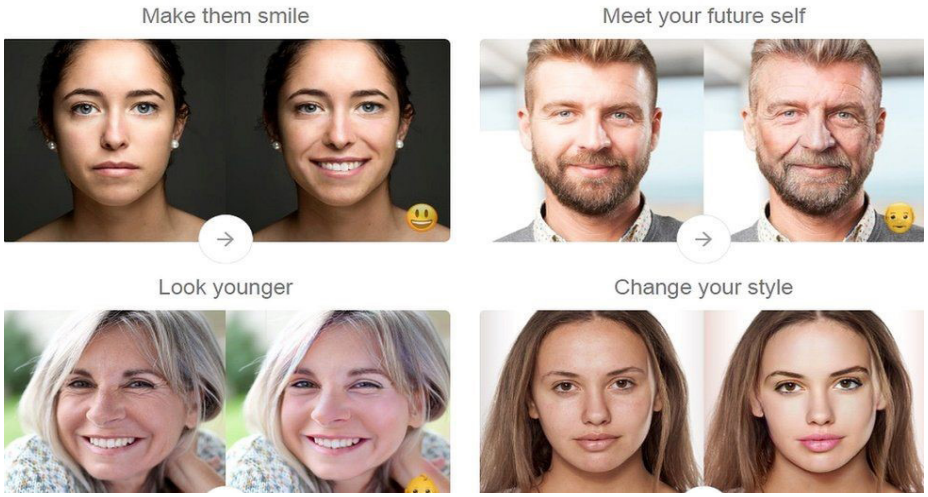
Durante los últimos años muchas de las aplicaciones móvil que se han popularizado nos permiten “jugar” con nuestras imágenes y vídeos, gracias a los avances en las técnicas de procesado de los mismos, y a las capacidades de inteligencia artificial.

Con estas aplicaciones podemos realizar retoques fotográficos (en algunos casos muy avanzados) sin necesidad de poseer conocimientos de fotografía o de procesado digital de imágenes. Simplemente es necesario disponer de un *smartphone* y descargar una de las múltiples aplicaciones disponibles. Un ejemplo, que muy posiblemente nos viene a la mente, son los filtros que pueden aplicarse en *Snapchat* o *Instagram*.

Algunos de estos retoques se limitan a modificar algunas propiedades de la imagen tales como el brillo, la saturación de color o el contraste, mientras que filtros más complejos nos permiten transformar radicalmente nuestras fotografías e incluso, en algunos casos, modificar nuestro rostro y apariencia por completo.

Quizá, las aplicaciones que más han proliferado son aquellas que nos permiten intercambiar nuestra cara con la de otras personas. Este proceso, conocido en inglés como *face swap*, pese a no ser algo nuevo, se ha democratizado al existir la posibilidad de realizarlo desde nuestros teléfonos con tan solo una fotografía en la que aparezca nuestro rostro.

Con algunas de estas aplicaciones podemos intercambiar las caras de dos personas que aparezcan en una foto, mientras que otras, algo más complejas y actuales, nos permiten sustituir la cara de un cantante o famoso en un vídeo por nuestro propio semblante.



Aplicaciones con *FaceSwap*. [1]

Otras, en cambio, nos permiten transformar nuestro rostro para cambiar nuestra apariencia por completo. Un ejemplo reciente, y que trajo mucha controversia[1], es *FaceApp*, con la que podemos recrear versiones femeninas o masculinas de nosotros mismos, o incluso ver nuestra apariencia con algunos años de más, o de menos. La controversia surgió porque no estaba claro en qué condiciones de seguridad almacenaban los datos, si los compartían con terceros o si los utilizaban para fines no declarados.

La mayoría de aplicaciones y herramientas que nos permiten realizar *face swap* se basan en algoritmos y modelos de inteligencia artificial. Al uso de técnicas de inteligencia artificial que hacen posible la modificación de los rasgos del rostro de una persona para hacerla pasar por otra se le conoce como *deepfake*. De hecho, y debido a los avances en esta área, también puede considerarse *deepfake* a la modificación de los movimientos de dicha persona, o incluso la generación de voz artificial. El término *deepfake* tiene su origen en el inglés y es un acrónimo formado por las palabras *deep* (de *deep learning*, o aprendizaje profundo) y *fake* (falso o falsificación).

La tecnología evoluciona cada vez más y más rápido, y los avances en inteligencia artificial no son una excepción. Esto muchas veces causa que no sea posible legislar a una velocidad que permita contemplar los posibles escenarios que ofrece la aplicación de nuevos desarrollos e incluso, en algunos casos, no existirá consenso sobre si está aceptado o no su uso. El *deepfake*, como no podía ser de otro modo, no está exento de controversia.

El avance en las técnicas y algoritmos de inteligencia artificial encargadas de recrear e intercambiar rostros ha permitido que los resultados obtenidos estén cada vez más logrados. Por poner un ejemplo del potencial de estos modelos, ¿se atrevería el lector a determinar si la siguiente imagen es real o, en cambio, ha sido generada por una máquina?

Imagen generada por ordenador de una persona que no existe [2]

Este tipo de técnicas, además de emplearse en multitud de sectores (como el cine o la publicidad), en muchas ocasiones son empleadas, desgraciadamente, en la generación de *fake* news o el chantaje a otras personas.

En el presente informe se introducirán algunas de las aplicaciones en las que más se está utilizando el *deepfake* y aquellas en las que su potencial uso tendría grandes ventajas. Para entender con algo de detalle cómo funcionan estas técnicas, se incluyen algunos fundamentos de los tipos de redes neuronales más utilizadas. Para finalizar, se discutirán los problemas que el *deepfake* conlleva, así como las implicaciones éticas y legales que pueden surgir en distintos sectores y/o casos de uso al generar datos de manera sintética.

Capítulo 2

Aplicaciones *DeepFake*

A continuación, se enumeran algunos de los principales ámbitos o sectores en los que este tipo de técnicas se aplican hoy en día o dónde podrían tener impacto directo en sus procesos productivos.

Redes sociales

El *face swap* ha tenido su mayor auge, a lo largo de los últimos 3 años, gracias a la aparición de múltiples aplicaciones para nuestros teléfonos móviles, así como a la mejora en la calidad de las cámaras, en la potencia de los procesadores y en las técnicas de procesado de imagen, nos permite realizar múltiples transformaciones en las fotos que compartimos en nuestra redes social.

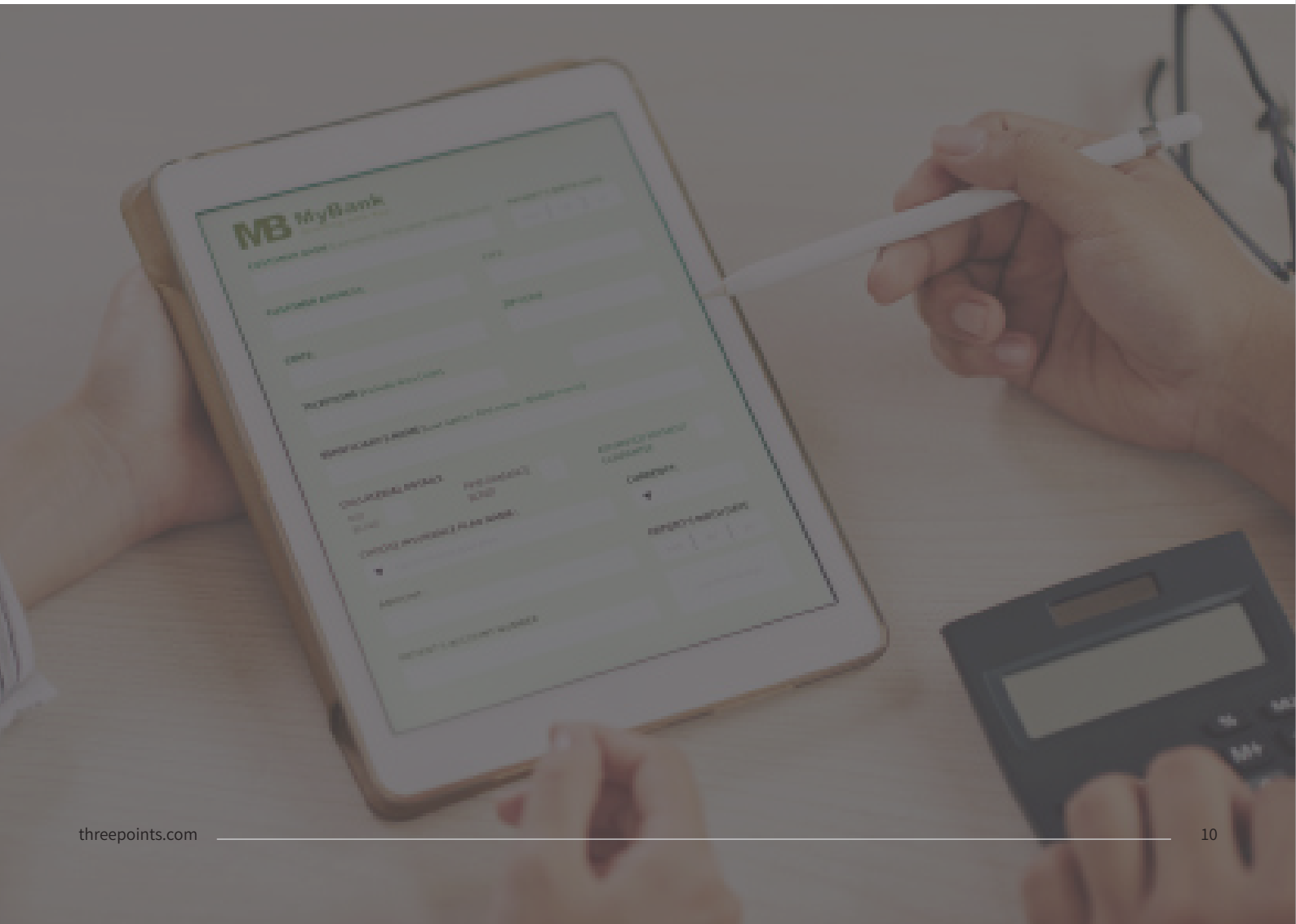
El objetivo de estas transformaciones puede ser distinto, por ejemplo, incrementar la calidad de la imagen mediante pequeños cambios en el color, brillo o contraste, o simplemente ocio. En el caso del *deepfake* usado en redes sociales es evidente que la principal aplicación es el *face swap*.

Lo habitual es que cuando descargamos e instalamos una aplicación debamos aceptar algunos permisos para que pueda funcionar correctamente. En ese momento, lo que hacemos es dar acceso a la aplicación a diferentes servicios de nuestro teléfono: cámara, geolocalización, micrófono… En muchas ocasiones, lo más común es que aceptemos sin percatarnos de lo que realmente estamos permitiendo. Esta es una mala práctica porque perdemos la noción de a qué posibles fuentes de datos pueden estar teniendo acceso terceras personas.

Pero realmente, lo más importante no es a qué fuentes de información tienen acceso las aplicaciones que instalamos, sino qué uso harán de esta información y si los datos que recaben cumplirán con el Reglamento General de Protección de Datos.

Es importante que cuando instalamos una aplicación nos detengamos a analizar los motivos por los que nos solicita acceso a información, y garanticemos que el uso que hará de la misma es completamente lícito. En cuanto al *deepfake*, al utilizar imágenes en las que aparecen nuestros rostros, es importante que nos aseguremos de todo esto con mucho mayor motivo.

Es decir, no debemos acudir a utilizar rápidamente este tipo de aplicaciones solo porque los resultados sean muy realistas o incluso divertidos. Es importante tener presente que cada vez es más habitual emplear factores biométricos en sistemas de autenticación (como nuestra huella, nuestra voz o nuestro rostro), por lo que compartir imágenes de nuestra cara con cualquiera no es una buena idea.



Cine

Por muchos seguramente sea conocido cómo en la película *"Rogue One"*, de la afamada saga cinematográfica *Star Wars*, se recreó digitalmente a algunos personajes. En concreto, pudimos ver de nuevo a la Princesa Leia y a Grand Moff Tarkin en pantalla.

Para aquellos que no les sea tan familiar este caso, por poner en contexto, la película se enmarca temporalmente entre los episodios III y IV. Esto implica que los actores que interpretaban a los personajes en 1977 hoy en día no podrían volver a hacerlo al haber envejecido o incluso en algunos casos, haber fallecido.

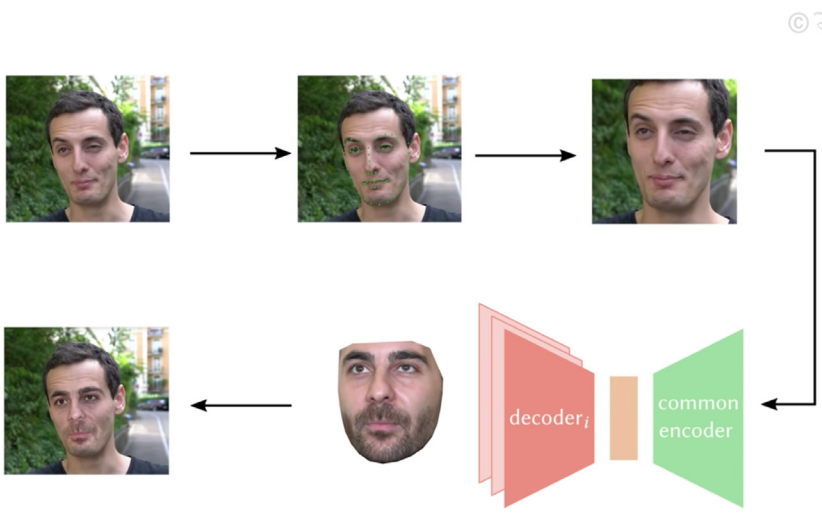
Lo que se decidió fue recrear los personajes mencionados anteriormente mediante CGI, siglas que se corresponden a Computer Generated Imagery, o, en castellano, Imágenes Generadas por Ordenador.

A la izquierda, personaje de Leia recreado digitalmente en Rogue One.
A la derecha, la actriz original en 1977.[4]

Existen otros muchos ejemplos del uso de CGI en el cine, no solo recreando personajes, sino también entornos o efectos especiales.

Pese a que las técnicas de CGI actuales no necesariamente hacen uso de redes neuronales artificiales, es evidente que se trata de un campo donde la aplicación del *deepfake* puede tener un gran potencial. Más adelante veremos cómo algunos tipos de redes neuronales son capaces de sustituir caras entre personas o recrear paisajes.

Actualmente, la aplicación de técnicas de *deep learning* en el cine es un área que se encuentra en desarrollo continuo con avances bastante prometedores. La empresa Disney, por ejemplo, ha desarrollado sus propios algoritmos[5] para el intercambio automático de rostros en alta resolución mediante técnicas de *deepfake*. Su objetivo es el de implementar estos avances en la producción de series y de películas suponiendo un ahorro significativo en costes operacionales y de producción.



Faceswap con el algoritmo de Disney.[5]

Además, un posible caso de uso que surge de la capacidad de poder actuar sobre el rostro de una persona en función de los gestos de una segunda, haría posible realizar doblajes mucho más realistas. En lugar de únicamente superponer la voz del doblador a la actuación durante el doblaje, el actor que es doblado podría realizar el mismo movimiento de labios que realiza el doblador de aplicar este tipo de técnicas.

Otro ámbito dónde el uso de técnicas de *deepfake* es bastante habitual, aunque en este caso es muy controvertido, es en el sector de la pornografía. De hecho, según el informe publicado por la empresa de ciberseguridad *Deeptrace*[6], se estima que alrededor del 96 % de todos los *deepfake*s que pueden encontrarse en internet tienen que ver con contenido pornográfico.

Publicidad

El sector de la publicidad es uno de los que más rápido evoluciona y se adapta a las nuevas tendencias. Sin ir más lejos, en muchas ocasiones, serán los anuncios y campañas de publicidad los propios generadores y potenciadores de estas tendencias.

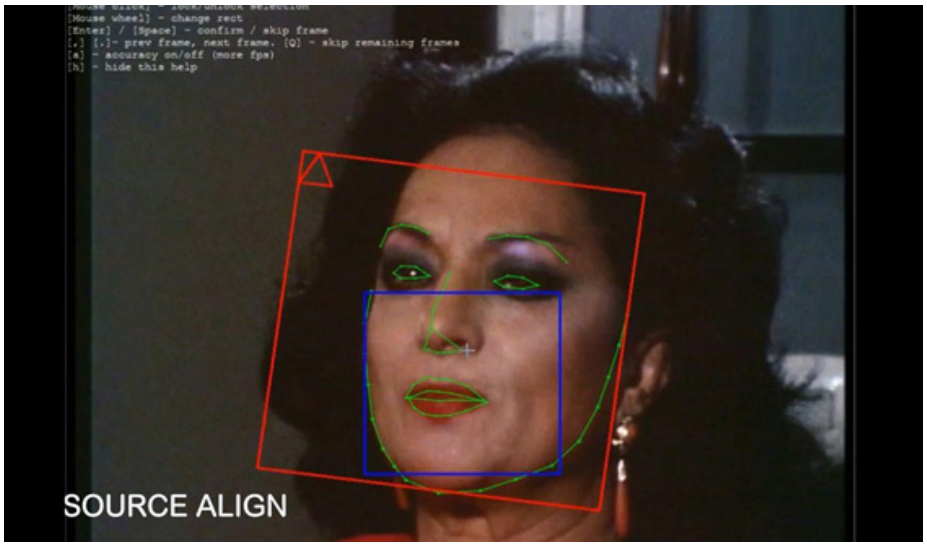
La publicidad es, después de todo, uno de los grandes early adopters de las nuevas tecnologías. Podemos pensar en el inmenso número de capacidades de inteligencia artificial del que se hace uso en este sector: recomendaciones personalizadas en aplicaciones -o portales de e-commerce- basadas en nuestros hábitos, modelos de clustering para segmentación de clientes, sistemas que predicen nuevas tendencias de consumo...

En el caso del *deepfake* no es ni mucho menos diferente. Son multitud de campañas de publicidad de diversas empresas en las que se han aplicado estas técnicas para recrear artificialmente la voz o la cara (o ambas) de las personas que aparecen en pantalla.

Por ejemplo, en 2019 la organización sin fines de lucro "*Malaria No More*", cuyo objetivo es detener la mortalidad de la malaria, hizo uso del *deepfake* en su campaña. En el anuncio[7], el exfutbolista inglés David Beckham lanza una petición para detener la enfermedad en 9 idiomas. A excepción del inglés, los ocho idiomas restantes son recreados de manera artificial en diferentes tonos. Durante todo el discurso David Beckham mueve los labios como si realmente él estuviese emitiendo las palabras en los distintos idiomas.

Spot de Hindi con David Beckham. [7]

Quizá, uno de los ejemplos que más repercusión ha tenido recientemente (al menos en el mercado español) ha sido la última campaña de Cruzcampo[8], empresa cervecera española. En ella, la empresa hace uso de esta tecnología "reviviendo" a la artista Lola Flores, fallecida en 1995, recreando artificialmente su rostro y su voz en un vídeo en el que aparece dando un discurso. En este vídeo se ilustran todos los pasos llevados a cabo durante la creación de este *deepfake*.



Making off del spot publicitario de Cruzcampo.[8]

La campaña ha reabierto el debate sobre estas tecnologías. Por un lado, hay quien halaga la inventiva y la técnica empleada, mientras que otros ponen en duda los límites éticos y legales de este tipo de prácticas.

En este sentido, es responsabilidad de la empresa anunciante usar este tipo de técnicas de una manera responsable y transparente.

Agentes conversacionales

Pese a que todos los ejemplos anteriores se centraban principalmente en la aplicación del *deepfake* en imágenes y/o vídeos, estos modelos pueden aplicarse también para síntesis de voz. En estos casos, el modelo aprendería la manera de hablar de una persona, siendo capaz de reproducir palabras o frases imitando su voz (síntesis del habla).

Esto permitiría enriquecer la experiencia de los usuarios con los asistentes virtuales al poder comunicarse con dispositivos que reproduzcan voces que les sean familiares. Por contra, como se detalla más adelante, este tipo de técnicas haría posible que estafas realizadas por teléfono fuesen más difícil de ser detectadas al ser las suplantaciones de identidad más realistas.

Capítulo 3

¿Cómo funcionan los *DeepFake?*

Hasta ahora hemos visto algunos de los sectores en los que el *deepfake* puede aplicarse con diferentes objetivos. Pero, ¿cómo funcionan realmente este tipo de técnicas? A continuación, se presentan algunos de los modelos que hacen esto posible.

Pero, antes de profundizar en ellos, es importante introducir algunos conceptos básicos sobre inteligencia artificial y redes neuronales.

Deep Learning

En el amplio mundo de la inteligencia artificial, el aprendizaje profundo, o como se conoce en inglés, *deep learning*, es una rama del machine *learning* basada en el uso de estructuras matemáticas que se inspiran en el funcionamiento del cerebro humano. A este tipo de estructuras se las denomina redes neuronales artificiales.

Estas redes neuronales artificiales están compuestas por unidades que se agrupan entre sí en una estructura de capas. Estas unidades, conocidas también como neuronas, tienen su origen en modelos matemáticos que tratan de modelar las neuronas que conforman nuestro cerebro.

La manera en la que se configuran las conexiones entre neuronas y capas define la arquitectura o topología de la red. Existen multitud de arquitecturas diferentes [9], siendo empleadas unos tipos u otros en función del caso de uso.

Por citar algunos ejemplos, las redes neuronales recurrentes[10] son especialmente útiles cuando trabajamos con datos donde la componente temporal es importante (análisis de series temporales, reconocimiento del habla, tareas de procesamiento del lenguaje natural…). Por otro lado, las redes neuronales convolucionales [11], han demostrado su efectividad trabajando en tareas tales como la clasificación de imágenes, la detección de objetos, o el reconocimiento facial.

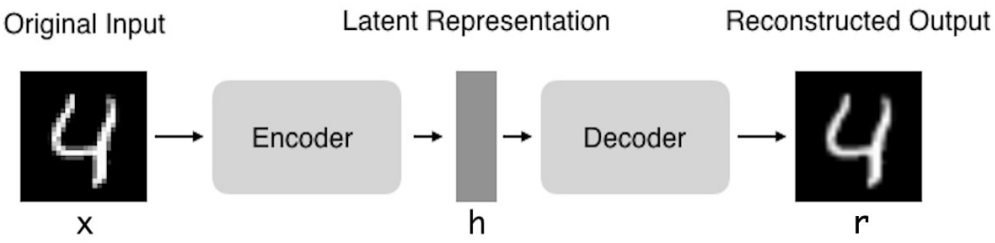
De entre todos los tipos de redes neuronales existe uno en concreto llamado autoencoder[12] que es una de las principales arquitecturas utilizadas en la creación de *deepfake*s.

Autoencoders

Este tipo de redes se emplean para aprender las codificaciones de los datos con los que se trabaje. En concreto, un autoencoder es un tipo de red neuronal que aprende cómo comprimir y codificar los datos a su entrada para, posteriormente, reconstruir los datos a partir de la versión comprimida y codificada de los datos de entrada, de forma que los datos a la salida sean lo más parecidos posible a los de la entrada. Durante el entrenamiento, la red aprende la mejor manera de reducir la dimensionalidad de los datos tratando de ignorar el ruido en los datos a la entrada.

La arquitectura típica de un autoencoder está formada por dos componentes:

- Encoder: el modelo aprende cómo reducir la dimensionalidad de los datos de entrada codificando la información en un espacio de variables latentes.
- Decoder: el modelo aprende cómo reconstruir los datos originales en base a la versión codificada de los datos de entrada.

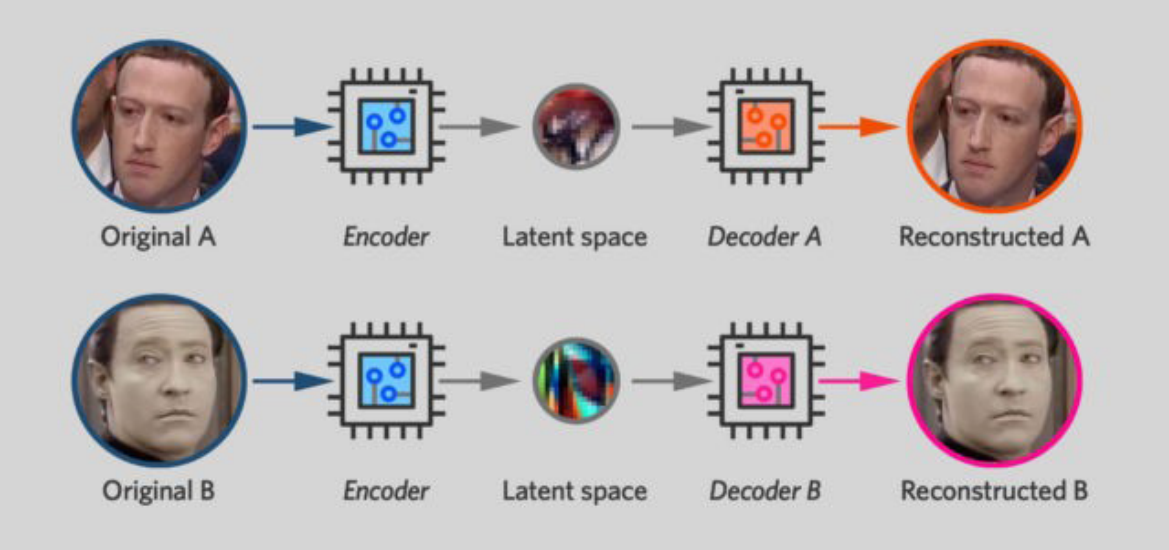


Arquitectura genérica de un autoencoder.[13]

A priori podría pensarse que entrenar una red para que devuelva a su salida una versión muy similar de los datos a su entrada no tiene mucha utilidad. Nada más lejos de la realidad, veamos a continuación por qué.

Los autoencoders, además de ser ampliamente utilizados en diversas tareas (como la detección de anomalías), tienen una particularidad que los convierten en modelos de gran utilidad a la hora de realizar *face swapp*ing. Al trabajar con imágenes con caras, el espacio de variables latente codificará información característica de los rostros como expresiones, dirección de la cara, posición de las cejas...

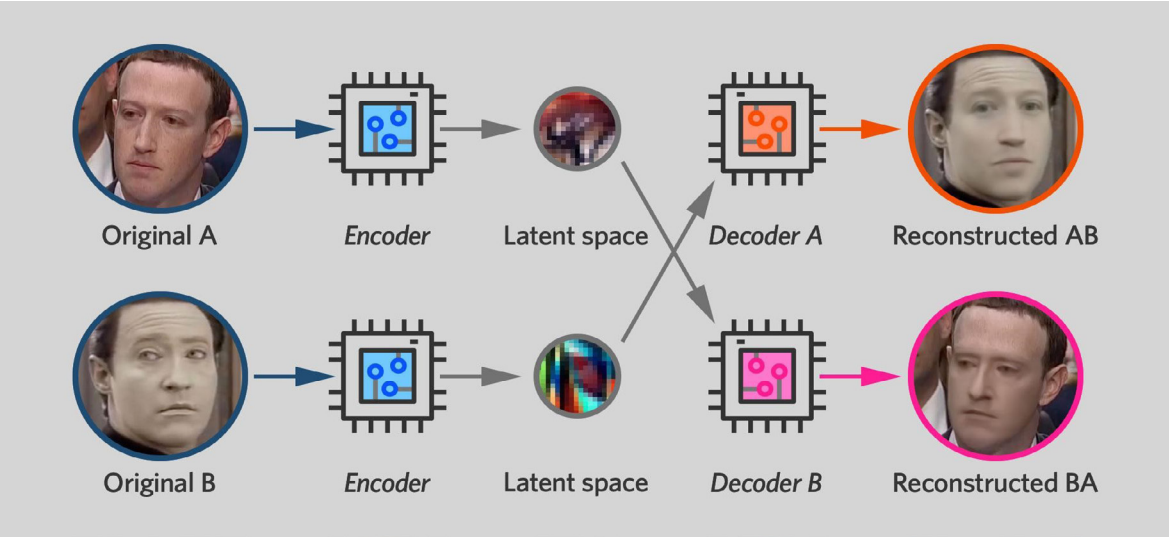
Imaginemos que queremos desarrollar un sistema que intercambie las caras dadas dos fotografías de personas que utilice autoencoders. Para ello, entrenaremos dos autoencoders, cada uno de ellos con fotografías de una persona distinta.



Entrenamiento de dos autoencoders con dos fotografías de dos personas distintas.[14]

Los autoencoders, si el entrenamiento se realiza correctamente, devolverán a su salida versiones muy similares a los datos de entrada. Si el entrenamiento de ambas redes se realiza en paralelo sin interferir entre ellas, cada autoencoder tendrá su propio espacio de variables latentes con características específicas de cada persona por separado.

Para solucionar esto, se fuerza que ambos autoencoders utilicen el mismo encoder. De esta forma, las características de ambas personas son codificadas en el mismo espacio de variables latentes. En cambio, a la hora de reconstruir la imagen, se emplean dos decoders distintos a la salida. Es decir, ambos autoencoders comparten el encoder, pero cada uno tiene su propio decoder. Durante el entrenamiento, cada encoder tratará de recrear solo las fotografías de una persona.



Autoencoders entrenados con los decoders intercambiados para realizar *face swapping*.[14]

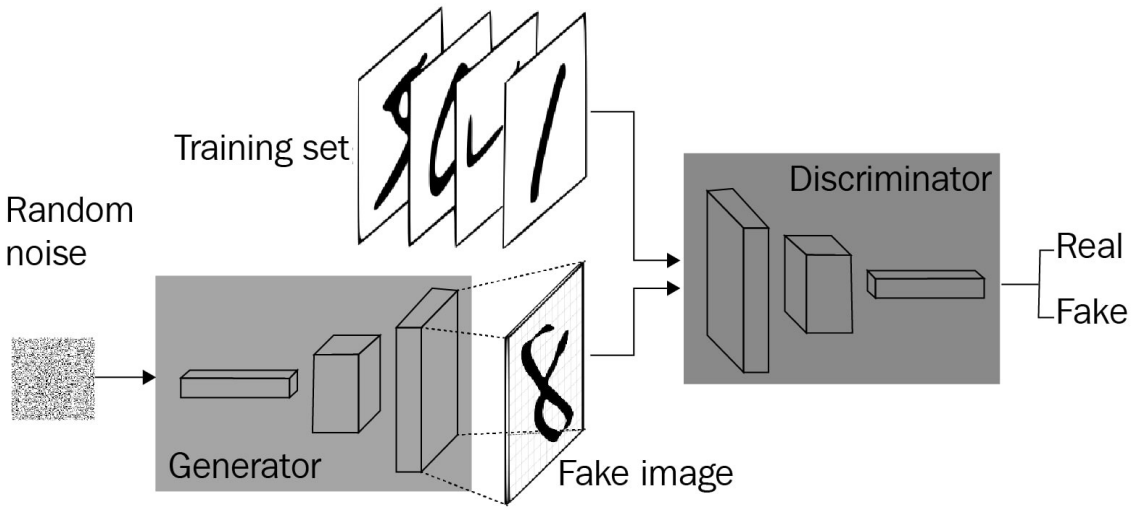
Una vez el proceso de entrenamiento ha finalizado, únicamente debemos intercambiar los decoders entre autoencoders.

Tomando como referencia la figura anterior, el autoencoder superior codificaría una fotografía de Mark Zuckerberg, pero a la hora de recrear la imagen se utilizaría el decoder del otro autoencoder. El resultado será una cara con la forma y expresión original de Zuckerberg, pero con algunos rasgos de la otra persona.

Generative Adversarial Networks

Otro tipo de redes neuronales que han demostrado ser excelentes a la hora de generar datos sintéticos son las llamadas Generative Adversarial Networks (GAN). Las redes GAN, propuestas por Ian Goodfellow en 2014[15] son un tipo de modelo de *deep learning* formadas por un sistema de dos redes neuronales que compiten entre sí. Permiten la generación de datos sintéticos y se emplean normalmente en la generación de imágenes.

Ambas redes, también conocidas como generador y discriminador, tienen objetivos distintos. Mientras que el generador construirá muestras de datos de entrenamiento falsas, el discriminador deberá determinar si las muestras son reales o no comparándolas con un conjunto de datos de entrenamiento. Ambas redes competirán entre sí durante el entrenamiento. Si el entrenamiento se realiza correctamente, el sistema generador será capaz de generar muestras cada vez más realistas.



Arquitectura general de una GAN.[16]

La evolución de este tipo de redes a la hora de generar rostros ha sido exponencial durante los últimos años. Los resultados, que en un inicio aparentaban ser prometedores, se han convertido como se aprecia en los últimos experimentos en recreaciones que fácilmente podrían engañar a cualquier persona.

Ejemplo de la evolución de las GAN durante los últimos años.[17]

Observando los resultados obtenidos recientemente puede entenderse porqué estas redes han sido una de las grandes revoluciones en el ámbito de la inteligencia artificial. Los siguientes rostros se han extraído de ThisPersonDoesNotExist.com[2], página web que cada vez que se refresca muestra un rostro generado artificialmente con una GAN entrenada con un conjunto de datos de caras enorme. El realismo de los rostros habla por sí solo.

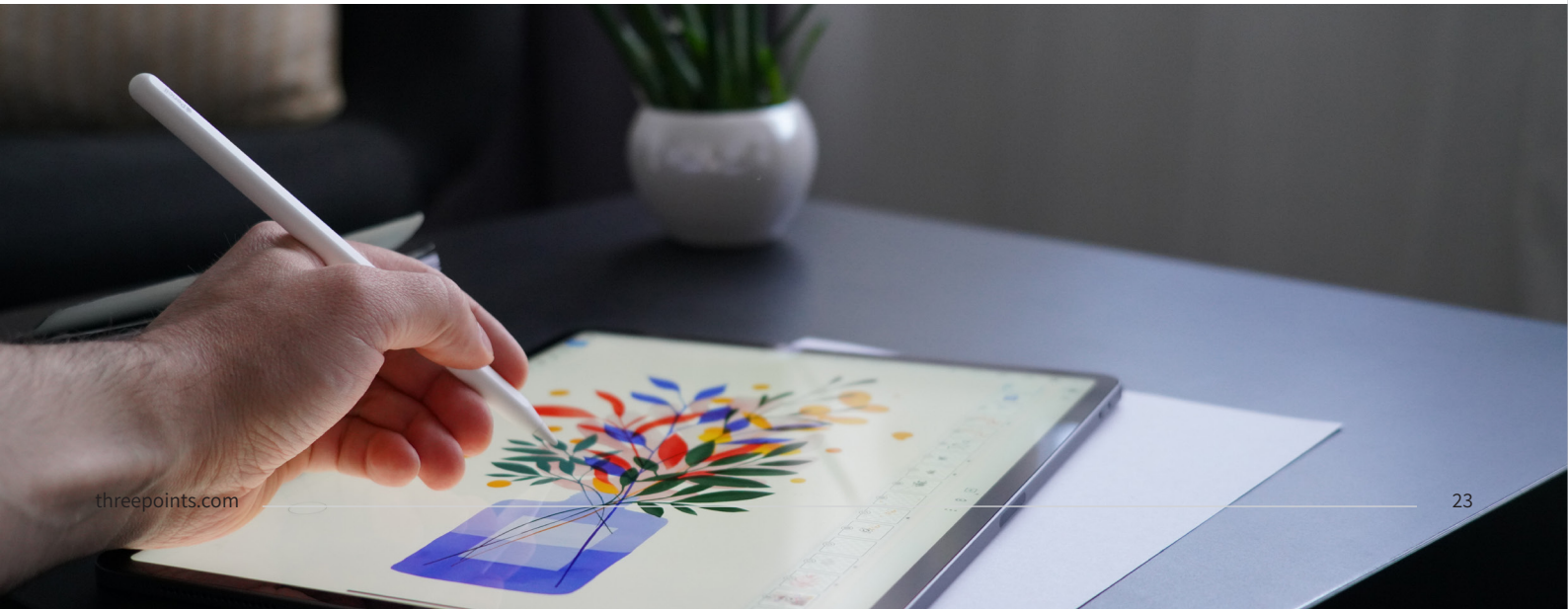
Ninguna de las caras es real. Todas han sido generadas por redes GAN. [22]

Neural Style Transfer

En *deep learning* existe una técnica de optimización de redes neuronales conocida como transferencia de estilo neuronal (o neural style transfer, en inglés) que permite combinar dos imágenes. Una de las imágenes se emplea como "contenido" y la otra como "referencia de estilo" (como las obras de un pintor determinado). Al combinarlas, el resultado es la imagen "contenido" aparentando tener el estilo de la imagen de referencia. Si tomamos como referencia el estilo de un pintor, el resultado podría verse como que la imagen que escogemos como contenido ha sido pintada por el artista escogido.

Transferencia de estilo aplicado sobre una fotografía tomando como referencia obras de distintos artistas.[20]

Esta técnica puede ser aplicada también a la generación de rostros. En este caso, se eligen dos fotografías en las que aparezca el rostro de una persona en cada una de ellas. El rostro de la imagen "contenido" copiará los rasgos de la imagen "referencia".



Veamos un ejemplo de esto con las siguientes imágenes generadas por modelos desarrollados por investigadores de Nvidia[22].

Transferencia de estilo aplicado sobre una fotografía tomando como referencia obras de distintos artistas.[22]

Las caras de la fila superior (empezando por la izquierda en el hombre con gafas de sol) se corresponden con las imágenes “contenido”. En cambio, las de la primera columna (comenzando desde arriba en la chica morena de pelo largo que sonríe) son aquellas imágenes de las cuales se extraerán los rasgos.

Puede verse como la red es capaz de extraer y combinar rasgos como el color de la piel y del pelo en los rostros originales, dando como resultado caras que bien podrían corresponderse con personas reales completamente nuevas.

Capítulo 4

Implicaciones éticas y legales

Como se ha mostrado, los avances en la generación de rostros sintéticos y la manipulación de vídeo y voz son abrumadores. Sin ir más lejos, con los resultados actuales, en muchos casos puede ser muy difícil, sino imposible, determinar si una imagen de una persona ha sido tomada por una cámara o si se trata de la salida de una GAN.

Esto nos lleva a pensar que posiblemente, no dentro de mucho, la diferencia entre la realidad y la ficción en los vídeos que veamos en televisión o en internet será apenas apreciable.

Existen diversas situaciones que hacen que el uso del *deepfake* puede conllevar problemas o, como poco, situaciones que desde el plano ético y legal aún no están del todo claras.

Fake news

Pese a que hoy en día estamos más conectados que nunca, el índice de penetración a internet de la población a lo largo del globo no para de crecer, y disponemos de multitud de fuentes de información (así como capacidades que nos permiten leer cualquier artículo en cualquier idioma gracias a la traducción automática), uno de los grandes problemas del siglo XXI es, irónicamente, la desinformación.

Compañías como *Facebook*, *Google* o Twitter tratan de luchar contra la generación de este tipo de contenido de diversas maneras. Y, pese a los esfuerzos de estos gigantes de internet, destinados a entrenar modelos cada vez más complejos que sean capaces de clasificar noticias como potencial contenido falso, o a desarrollar herramientas que permitan a sus usuarios reportar *fake* news, el problema no parece tener fácil solución.

El caso de las *fake* news es quizá un ejemplo paradigmático de cómo los *deepfake*s pueden ayudar a propagar las *fake* news en la población ya que con herramientas -en muchos casos- al alcance de cualquiera, es posible crear, por ejemplo, un vídeo en el que aparezca una persona emitiendo un mensaje que nunca existió.

Un ejemplo muy conocido es el de Obama declarando en un vídeo (falso, por supuesto) que “El presidente Trump es un total y completo idiota”.[24]

El impacto que este tipo de manipulaciones pueden tener es enorme. Pensemos en qué podría ocurrir si en un *deepfake* aparece un CEO de una gran empresa o un primer ministro de un país que hace que se hunda su reputación o diciendo alguna mentira. Si el vídeo es realista muy posiblemente se propagará mucho más rápido que la verdad sobre la autenticidad del mismo. ¿Cuál podría ser el impacto en su imagen, en los mercados o en la política?

Fragmento del vídeo en el que se muestra a la persona que “controla” a Obama durante el *deepfake*. [24]

Scam

Scam es un término inglés que hace referencia a las estafas o fraudes realizados mediante medios electrónicos y/o internet. Este tipo de tretas suelen ser muy comunes en medios como el correo electrónico, los SMS, o redes sociales. Es en estas últimas donde el uso de *deepfake*s toma mayor relevancia al existir la posibilidad de poder suplantar la imagen de una persona e, incluso, su voz.

Un caso muy sonado fue el de un CEO de una empresa de energía basada en Reino Unido. Según se menciona en el artículo que lo reporta[25], la persona pensaba que estaba al teléfono con su superior, el CEO de la empresa matriz. Durante la conversación el superior le pedía que realizase un pago de más de 200.000 $. En realidad, la voz del CEO de la empresa matriz fue recreada con técnicas de inteligencia artificial, por lo que es fue un claro caso de scam debido a un *deepfake*.

Recrear a personas ya fallecidas

Muchos de los casos en los que el *deepfake* ha demostrado ser de gran utilidad es a la hora de recrear personas o personajes en series o películas.

Pero, desde el punto de vista legal, ¿qué ocurre si estas personas ya han fallecido? La ley es clara en qué está permitido a la hora de usar imágenes o fragmentos de vídeos en los que aparezca una persona ya fallecida. En cambio, y al igual que ocurre con otras tecnologías, la ley no especifica qué ocurre con los *deepfake*s. En principio, y dado que el derecho de imagen de una persona desaparece cuando ésta fallece, serán los herederos o familiares más cercanos los que deberían autorizar el uso de la imagen del fallecido en un *deepfake*. Esto dependerá, lógicamente, de la legislación de cada país o estado.

Pese a que desde el plano legal pueden estar definidos los límites, desde el plano de la ética existe mucha controversia. Sin ir más lejos, los Simpson (como en otras muchas ocasiones), ya ironizaban con este tema [26].

Por otro lado, también hay quien reclama que el espectador debe conocer si las personas que aparecen en un spot publicitario, así como sus actos y voces, han sido recreados por ordenador.

Capítulo 5

Conclusiones

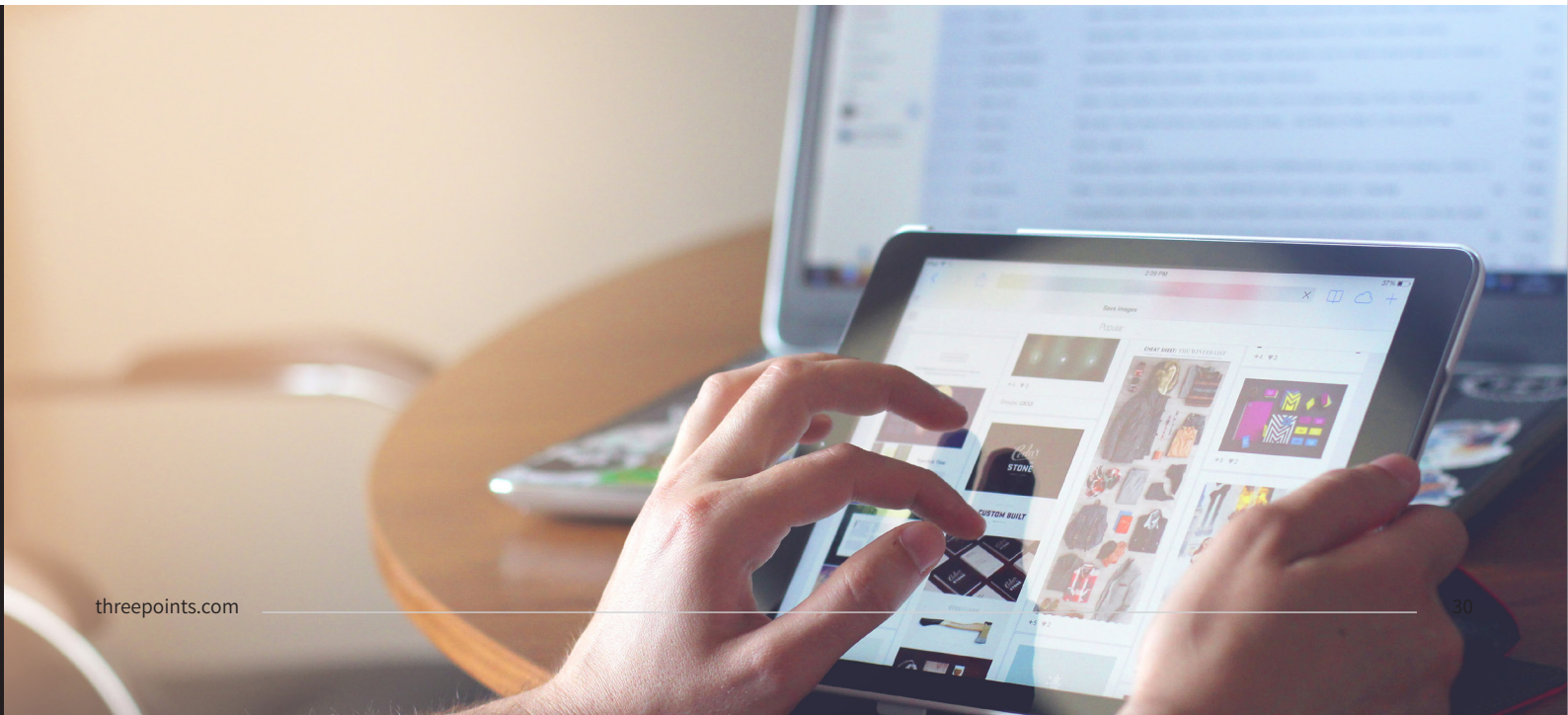
Estamos ante un escenario en el que las diferentes disciplinas dentro del inmenso espectro que ofrece la inteligencia artificial evolucionan a un ritmo acelerado. No solo día a día aparecen nuevos y prometedores avances, sino que incluso se descubren técnicas que pueden ser aplicadas a sectores que hasta el momento no se habían planteado.

Un ejemplo evidente son los *deepfake*s y como gracias a ellos muchos sectores, como puede ser la industria cinematográfica, pueden verse beneficiados gracias a los impresionantes resultados obtenidos en la generación de rostros y voces de manera artificial.

Además de las aplicaciones directas de este tipo de algoritmos, gracias al desarrollo de modelos matemáticos cada vez más complejos que han permitido que los *deepfake*s sean una realidad, aparecerán nuevos modelos que serán capaces de realizar tareas que a día de hoy no imaginamos que puedan ser posibles.

Desgraciadamente, y como ocurre con todo gran avance en la ciencia y tecnología, tiene su parte negativa. Los *deepfake*s, como hemos visto, pueden ser utilizados para diferentes actos delictivos como pueden ser la suplantación de identidad o la creación de vídeos en los que aparece una persona emitiendo un discurso que, aun pareciendo ser reales, son completamente falsos. El coste reputacional que esto puede acarrear a los protagonistas de los *deepfake*s, así como el impacto directo que puede tener en los mercados o en la política, son demasiado elevados como para no considerar un gran riesgo es posible uso ilegítimo que se pueda hacer con los *deepfake*s. Sin duda, lo que parece ser necesario es una regulación rápida de las normas que permitan determinar cuáles son los límites a la hora de generar *deepfake*s.

Por otro lado, como usuarios de redes sociales y consumidores de medios digitales y plataformas de vídeo, es importante mantener un pensamiento crítico que nos haga recordar que no todo lo que vemos o leemos tiene por qué ser verdad siempre. Es fundamental acudir a fuentes de información que sean confiables y que incluyan referencias. Además, nuestra responsabilidad es doble. De detectar un *deepfake* que trata de hacer daño a una persona o de propagar una *fake* new, es fundamental que no colaboremos en la difusión de contenido *fake* y, si es posible, se reporte o denuncie a quien sea conveniente en cada circunstancia.



Referencias bibliográficas

1 Baraniuk, B. C. (2019, 17 de julio). Can you trust *FaceApp* with your *face*? BBC News. https://www.bbc.com/news/technology-49018103

2 This Person Does Not Exist. (2021). ThisPersonDoesNotExist. https://thispersondoesnotexist.com/

3 High-Resolution Neural *Face Swapp*ing for Visual Effects | Disney Research Studios. (2020, 29 junio). Disney Research Studios. https://studios.disneyresearch.com/2020/06/29/high-resolution-neural-*face-swapp*ing-for-visual-effects/

4 RANKED: The 23 best CGI-enhanced movies ever. (2016, 15 abril). Business Insider. https://www.businessinsider.com/best-cgi-movies-2016-4?international=true&r=US&IR=T#1-star-wars-1977-23

5 DisneyResearchHub. (2020, 29 junio). High Resolution Neural *Face Swapp*ing for Visual Effects. YouTube. https://www.youtube.com/watch?v=yji0t6KS7Qo

6 *Deep*trace (2019, 1 de octubre). The State of *Deepfake*s - Landscape, Threats, and Impact. https://regmedia.co.uk/2019/10/08/*deepfake*_report.pdf

7 Malaria Must Die. (2019, 9 abril). David Beckham speaks nine languages to launch Malaria Must Die Voice Petition. YouTube. https://www.youtube.com/watch?v=QiiSAvKJIHo

8 elDiario.es. (2021, 23 de enero). Los creadores del *deepfake* de Lola Flores: «Ya no se puede confiar en que todo lo que se ve en un vídeo es real». eldiario.es. https://www.eldiario.es/tecnologia/nadie-mejor-lola-flores-recordar-potencia-tecnologia-*deepfake*_1_7036932.html

9 Veen, F., & Leijnen, S. (2016). A mostly complete chart of Neural Networks. The Asimov Institute. https://www.asimovinstitute.org/neural-network-zoo/

10 Rumelhart, David E.; Hinton, Geoffrey E., y Williams, Ronald J. (1985). *Learning* internal representations by error propagation. Tech. rep. ICS 8504. San Diego, California: Institute for Cognitive Science, University of California. https://*app*s.dtic.mil/dtic/tr/fulltext/u2/a164453.pdf

11 LeCun, Y. (1989). Generalization and network design strategies. Tech. rep. CRG-TR-89-4, University of Toronto. http://yann.lecun.com/exdb/publis/pdf/lecun-89.pdf

12 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (pp. 499-523). MIT Press. https://www.*deeplearning*book.org/contents/autoencoders.html

13 *DeepLearning*Italia (2018, 8 de mayo). Introducción al autoencoder. *deeplearning*italia.com. https://www.*deeplearning*italia.com/introduzione-agli-autoencoder-2/

14 Ars Technica (2019, 16 de diciembre). I created my own *deepfake*—it took two weeks and cost $552. arstechnica.com. https://arstechnica.com/science/2019/12/how-i-created-a-*deepfake*-of-mark-zuckerberg-and-star-treks-data/

15 Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-farley, D., Ozair, S., Courville, A., y Bengio, Y. Generative Adversarial Nets. https://arxiv.org/pdf/1406.2661.pdf

16 Packt. Machine *Learning* Projects for Mobile *App*lications. Karthikeyan NG. https://subscription.packtpub.com/book/big_data_and_business_intelligence/9781788994590

17 TimMcCloud (2020, 2 de marzo). Comprehensive introduction to turing *learning* and gans. timmcloud.com. https://timmccloud.net/blog-comprehensive-introduction-to-turing-*learning*-and-gans-part-2/

18 neptune.ai (2020, 2 de abril). 6 GAN Architectures You Really Should Know. neptune.ai. https://neptune.ai/blog/6-gan-architectures

19 Karras, T., Laine, S., y Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. https://arxiv.org/pdf/1812.04948.pdf

20 Gatys, Leon A., Ecker, Alexander S. y Bethge, M. A Neural Algorithm of Artistic Style. https://arxiv.org/pdf/1508.06576.pdf

21 Brock, A., Donahue, J. y Simonyan, K. Large Scale GAN Training for High FIdelity Natural Image Synthesis. ICLR. https://arxiv.org/pdf/1809.11096.pdf

22 Vincent, J. (2018, 17 de diciembre). These *face*s show how far AI image generation has advanced in just four years. The Verge. https://www.theverge.com/2018/12/17/18144356/ai-image-generation-*fake*-*face*s-people-nvidia-generative-adversarial-networks-gans

23 Zhu, J., Park, T., Isola, P. y Efros, A (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. UC Berkeley. ICCV. https://junyanz.github.io/CycleGAN/

24 BuzzFeedVideo. (2018, 17 abril). You Won't Believe What Obama Says In This Video! YouTube. https://www.youtube.com/watch?v=cQ54GDm1eL0&feature=emb_logo

25 Stupp, C. (2019, 30 de agosto). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. WSJ. https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

26 Bart Zombie. (2018, 21 de febrero). The Simpsons - How to Get Ahead in Dead Vertising. YouTube. https://www.youtube.com/watch?v=AwLRCAU3dus&feature=youtu.be

THREEPOINTS
THE SCHOOL FOR DIGITAL BUSINESS

De:
Planeta Formación y Universidades