



El Rol de los Deepfakes en la Comunicación Política: Análisis de Regulaciones, Amenazas y Beneficios en la Era Digital (2019-2022)

Tutor: Rodolfo Lautaro Bruera

Tesista: Mateo Corrado

Licenciatura en Ciencia Política

Facultad de Derecho y Ciencia Política

12 de Julio de 2023

Índice

1. Cuestiones Metodológicas-Teorías	04
1.1. Justificación y/o Relevancia del Tema	04
1.2. Preguntas de la Investigación.....	07
1.2.1. Pregunta Central de Investigación	07
1.2.2. Preguntas Específicas de Investigación	07
1.3. Hipótesis	08
1.4. Objetivos	09
1.4.1. Objetivo General	09
1.4.2. Objetivo Específico.....	09
1.5. Antecedentes	09
1.6. Marco de Teórico	11
1.6.1. Inteligencia Artificial	12
1.6.1.1. Machine Learning	12
1.6.1.2. Deep Learning.....	12
1.6.1.3. Generative Adversarial Networks.....	12
1.6.1.4. Synthetic Media	13
1.6.1.5. Deepfakes.....	13
1.6.2. Amenazas de los deepfakes	13
1.6.2.1. Estafas y Chantajes	14
1.6.2.2. Erosión de la Verdad.....	14
1.6.2.3. Debilitamiento del Periodismo y los Medios de Comunicación.....	16
1.6.2.4. Erosión de las Instituciones	17

1.6.3. Beneficios de los Deepfakes	19
1.6.3.1. Entretenimiento.....	20
1.6.3.2. Protección de la Identidad.....	22
1.6.3.3. Investigaciones Médicas y Psicológicas	23
1.6.3.4. Educación.....	23
1.6.3.5. Usos Comerciales.....	24
2. Deepfakes, Inteligencia Artificial y el Impacto en la Comunicación Política.....	25
2.1. Creación y Detección de Deepfakes	25
2.2. La Inteligencia Artificial en la Comunicación Política.....	28
2.2.1. Carrera Armamentista por la Inteligencia Artificial	32
2.2.1.1. Rusia	32
2.2.1.2. Francia.....	32
2.3. Los Deepfakes en la Comunicación Política	33
2.3.1. Cambios en las Formas de Comunicación	34
2.3.2. Casos de Deepfakes en la Política	36
3. Regulaciones y Soluciones	37
3.1. Regulaciones	37
3.1.1. Europa	38
3.1.2. Estados Unidos.....	39
3.1.2.1. California (AB 730)	39
3.1.2.2. Texas (S.B. NO. 751).....	39
3.1.3. China	39
3.1.4. Redes Sociales	40

3.2.Soluciones	41
3.2.1. Soluciones Mediante Autenticación	41
3.2.2. Soluciones con Tecnología	41
3.2.3. Soluciones con la Blockchain	42
3.2.4. Otras Soluciones	43
4. Ética.....	43
5. Conclusión	44
6. Referencias Bibliográficas.....	47

Cuestiones Metodológicas-Teorías

Justificación y/o relevancia del tema

El *Synthetic Media* está cambiando la forma en la que producimos contenido audiovisual. Esta tecnología permite a los usuarios crear contenido artificial que imita al creado por humanos. Ajder & Glick (2021) definen el Synthetic Media como “contenido creado o modificado por una inteligencia artificial” (p.9)

Dentro de este amplio área encontramos el término *deepfake*, que se refiere al uso de técnicas complejas de Machine Learning (Deep Learning), para crear contenido artificial que aparenta ser real. Principalmente imágenes, videos o audios de personas diciendo o haciendo cosas que nunca hicieron o dijeron.

Esta tecnología fue vista por primera vez en el 2017 en un subforo de Reddit, y desde entonces ha crecido rápidamente. Estos metrajes, por lo general, requieren una gran cantidad de datos para entrenar modelos que generen contenido realista. Mientras que en el pasado, los deepfakes tenían señales evidentes de ser falsos, los recientes avances en la tecnología hacen que sea cada vez más difícil distinguirlos

Los deepfakes pueden ser vistos entonces como una amenaza o una oportunidad. El objetivo de esta investigación es explorar el emergente fenómeno de los deepfakes durante el lapso temporal 2019-2022, haciendo foco en entender por qué y cómo son creados para comprender sus impactos en la sociedad.

Existen diversos motivos por lo cual es importante estudiar el impacto de los deepfakes desde la Ciencia Política. Aunque por ahora los deepfakes no parecen una amenaza para la sociedad, existe el riesgo de que la desinformación política y la creación de contenido no consensuado pueda afectar la seguridad nacional, los procesos judiciales y el bienestar general.

Bellomo y Oszlak (2020) afirma que muchas de estas innovaciones crean importantes desafíos y amenazas, y “en todos ellos el Estado cumple un rol insustituible, tanto para promover la innovación tecnológica como para resguardar a sus ciudadanos de las propias acechanzas que ese desarrollo genera.” (p.29)

En el pasado hacer un video era más difícil y caro que crear un artículo de texto. Hoy los contenidos audiovisuales basados en Deep Learning, incluyendo a los deepfakes originales, requieren mucho menos tiempo, habilidades artísticas, conocimientos técnicos y recursos informáticos para producirlos (Millière, 2022, p.16)

Estamos en un tiempo en la cual está evolucionando la forma de consumir información. Primero, apareció Twitter, una red social basada principalmente en texto; luego surgió Facebook, que tenía mucho texto y algunas imágenes; más tarde Instagram, donde las fotos tomaron relevancia; hoy tenemos a TikTok, que se basa exclusivamente en video. La comunicación visual es ahora la norma, no la excepción.

Esta época está también marcada por el consumo de información rápida y estimulante como videos con subtítulos, tweets, anuncios publicitarios y mucho más. Esto hace a las personas más vulnerables a la influencia de la información recibida.

El estudio de Chesney y Citron (2019) hace un análisis completo de la legislación californiana sobre deepfakes. Esta investigación enumera una amplia cantidad de amenazas y beneficios de los deepfakes. Entre las principales amenazas que surgen de la aplicación de esta tecnología, los autores destacan las siguientes:

- El aumento del Liar's Dividend: al combinar el debilitamiento de la verdad con el debilitamiento de la confianza el autoritarismo va a ganar espacio, y así “los regímenes

autoritarios y los líderes con tendencias autoritarias se van a beneficiar cuando la verdad objetiva pierda su poder.” (Chesney & Citron, 2019, p.1786)

- El Debilitamiento del periodismo y los medios de comunicación: Cuando uno es consciente de que un video es artificial, nuestra percepción no es igual que si no estuviéramos al tanto de esto. Chesney y Citron (2019) dicen que, en consecuencia, “las organizaciones de noticias podrían dejar de reportar rápidamente algo como real por miedo a que la evidencia sobre estos termine por ser falsa.” (p.1784)
- La erosión de las Relaciones Internacionales: “las agencias de inteligencia de los gobiernos más poderosos están en una buena posición de tomar decisiones inteligentes sobre el peso que les dan a potenciales fakes.” (Chesney & Citron, 2019, p.1782)

A su vez, los autores también comentan que los deepfakes tienen aplicaciones beneficiosas, entre las que destacan los fines educativos y su utilización en las manifestaciones mediante la sátira. Por otro lado, la investigación de Tammekänd et al. (2020) los diferencia en cinco tipos de deepfakes.:

- Facial re-enactment deepfake: la manipulación de expresiones faciales de un objetivo en un video basado en el actor fuente
- Face swapping: una cara de alguien cambiada en otra cara objetivo.
- Full body deepfakes: incorpora las técnicas del face swapping y facial re-enactment pero extiende la tecnología de inteligencia artificial para ahora generar sintéticamente un cuerpo que se mueva.
- Audio deepfakes: se refiere a la tecnología de inteligencia artificial que es usada para generar sintéticamente palabras que imitan a un objetivo. Los deepfakes de audio pueden

resultar en una conversación tecnológica o mensaje de voz siendo generado sintéticamente e indescifrable para el oído humano.

- Deepfakes basados en humanos sintéticos: humanos generados sintéticamente con GANs y técnicas deepfakes pueden ser combinadas para generar deepfakes basados en humanos sintéticos.

El European Parliament et al. (2021) aportan un importante análisis sobre las regulaciones de deepfakes, discutiendo solo los principales marcos legales sobre el tema. Véase: marco regulador sobre la inteligencia artificial, regulación general de la protección de datos, ley de copyright, derechos de imagen, servicios digitales, servicios de medios audiovisuales, medidas contra la desinformación y resoluciones del European Parliament contra los deepfake.

Los autores alertan que “nuestros ecosistemas de información están ahora caracterizados por una velocidad aumentada, sobreinformación, pero también por noticias cortas y fragmentadas, como por ejemplo, los hilos de Twitter.” (European Parliament et al., 2021, p.22)

Preguntas de la Investigación

Pregunta Central de Investigación

¿Cuáles son las motivaciones de quienes utilizan los deepfakes en la comunicación política y cuáles fueron las implicaciones positivas y negativas de la utilización esta tecnología a nivel mundial en el lapso temporal 2019-2022?

Preguntas Específicas de Investigación.

- ¿Cómo fueron creados los deepfakes, que técnicas y métodos fueron utilizados durante el lapso 2019-2022?
- ¿Cuáles son los mecanismos de detección que se utilizaron para identificar deepfakes durante el lapso 2019-2022?

- ¿Por qué se crean deepfakes maliciosos durante el lapso 2019-2022?
- ¿Cuál es la línea que marca la diferencia entre desinformación intencional y la sátira?

Hipótesis

A partir de los objetivos planteados, el enfoque que se le dará a la investigación será descriptivo y exploratorio, porque se buscará definir el fenómeno de los deepfakes, teniendo en cuenta métodos de creación, detección y difusión, para así poder explorar los posibles usos positivos y negativos de esta tecnología, las motivaciones de quienes los crean y difunden, y las consecuencias de su uso en la comunicación política.

Los deepfakes son metrajes falsos creados a través de GANs (Generative Adversarial Networks), que pueden ser utilizados tanto para fines positivos y negativos. A través de estos videos se pueden crear declaraciones políticas artificiales, crear consciencia sobre una determinada temática y llevar mensajes a minorías.

Así como una GAN puede crear un deepfake, también lo puede detectar. Esto es un problema, pues hace que los deepfakes mejoren cada vez más su calidad y realismo, lo que genera una lucha constante entre la creación de mecanismos de detección y técnicas nuevas de creación de deepfakes.

Los creadores de deepfakes maliciosos tienen como principal objetivo difamar, extorsionar y/o chantajear a una persona u organización. En el campo de la política esto podría ser utilizado para crear declaraciones políticas falsas, crear escenarios que no son reales y generar un ambiente tóxico en el que abunda la desinformación.

Un productor de deepfakes se puede defender afirmando que el video es una sátira, sin embargo existe una línea muy fina entre la desinformación intencional y la sátira, que está marcada principalmente por la intención y el contexto en el que se exhibe el deepfake.

Objetivos

Objetivo General.

Explorar el emergente fenómeno de los deepfakes durante el lapso temporal 2019-2022, haciendo foco en entender por qué y cómo son creados para comprender sus impactos en la sociedad.

Objetivos Específicos.

- Investigar cómo es creado un deepfake, incluyendo los métodos técnicos utilizados y como circulan en las redes sociales.
- Explorar las motivaciones de quienes crean este tipo de contenido
- Examinar los potenciales usos positivos y negativos de los deepfakes en la comunicación política.
- Examinar el impacto de los deepfakes en la sociedad, haciendo principal foco en su potencial para debilitar la credibilidad de las noticias e información.

Antecedentes

El estudio de Pinhanez et al. (2022), en representación de IBM, ha demostrado que nos encontramos ante una nueva ola de desinformación y que ya no es un asunto sociocultural, sino de ciberseguridad. Esta segunda ola de desinformación causa daño a la gente a nivel personal, cosa que no sucedía en la primera.

En consecuencia, los mecanismos utilizados en la actualidad como la verificación manual de contenidos, los grupos fact-check humanos o la eliminación del contenido falso ya no son efectivos. Se propone entonces dejar de lado las campañas masivas y pasar a un modelo en el que se ayude a los individuos que tuvieron *engagement* con una noticia falsa y la creyó.

Además, Pinhanez et al. (2022) proponen la creación de un software que verifique la información: un ‘Fake Media protector’. Los autores consideran que los mecanismos fact-checking

manuales ya están obsoletos y es necesario automatizarlos. Ante el aumento de la dificultad para distinguir entre lo real y lo sintético proponen la creación de un ecosistema de ciberseguridad el cual se analice, detecte e identifique contenido falso automáticamente.

Esta fuente establece un precedente, pues examina el panorama general del ecosistema digital con respecto a la desinformación, concluyendo que la tecnología actual no es suficiente para combatirla. Los autores del texto argumentan que el único modo de erradicar la desinformación es a través del acceso a la información personal de los ciudadanos, para así lograr argumentos convincentes.

Por otro lado, el estudio de Meneses (2021), con un análisis sobre las elecciones del 2020 en Estados Unidos y los deepfakes, lo que se suponía que ocurriría a partir de la utilización de los deepfakes en la elección y lo que finalmente sucedió. El autor plantea que para amenazar a Estados Unidos hoy ya no se necesita armamento militar, sino que basta con la propagación de Fake News. Explica como un deepfake puede ser un arma en elecciones y cita casos previos como el de Gabón, Italia e India.

Además, coincide con Pinhanez et al. (2022) en que no hay leyes que regulen lo que pasa en internet, que existe una necesidad de crear sistemas fact-checking y que los mecanismos de detección no estaban y no están preparados para una elección.

Otro antecedente importante es el texto de Gregory (2021), director de WITNESS, una ONG que utiliza videos y tecnología para proteger y defender derechos humanos. El informe habla sobre la mis información durante el COVID-19, es decir, la modificación de información real con el fin de desinformar.

El texto hace principal foco en las nuevas formas de manipulación y como ya no existe el ‘ver para creer’. La principal preocupación de Gregory (2021) son los deepfakes como forma de

manipulación en testimonios, y hace principal foco en el caso de Myanmar, donde un presidente fue acusado de corrupción después de que se difundiera un deepfake suyo.

El texto de Gregory es fundamental para entender el rol que tendrán los periodistas en el futuro, y como estos pueden recuperar la confianza de la gente mediante la autenticación de información en la era digital de la posverdad.

Por último, el texto *Just Joking*: escrito por Ajder y Glick (2021) en representación de WITNESS y el MIT Open Documentary Lab, este informe trata principalmente a la utilización de deepfakes en la sátira, se plantean los límites de la sátira y se hace la pregunta de qué pasa cuando la gente vulnerable no entiende la sátira.

El texto afirma que los deepfakes son un arma de doble filo, y que algunos regímenes autoritarios los pueden llegar a utilizar a su favor. Sin embargo, también comenta sobre casos en los que este tipo de Synthetic Media es utilizado con fines positivos, como lo fue la campaña de Reino Unido contra la malaria o el documental *Welcome To Chechnya*.

Ajder y Glick (2021) dan sus opiniones sobre donde se marca la línea ética de la utilización de deepfakes y como este nuevo tipo de contenido puede llevar al aumento del Liar's Dividend. Además, hacen un importante estudio de las políticas que aplican Facebook, Twitter, TikTok y YouTube contra el contenido artificial.

Marco Teórico

La investigación se propone analizar el papel que desempeñaron los deepfakes en la comunicación política durante el período 2019-2022, haciendo énfasis en las regulaciones y soluciones propuestas durante ese período. Es importante destacar que, debido a la escasez de literatura específica en el campo de la ciencia política sobre esta temática en particular, se debió recurrir a

fuentes en inglés y explorar estudios provenientes de disciplinas conexas como las ciencias computacionales.

Inteligencia Artificial

La Inteligencia Artificial es la ciencia que intenta replicar la inteligencia y sus procesos a través de las computadoras. Girardi (2019) expande este concepto al afirmar que la I.A. “lleva consigo la posibilidad del autoaprendizaje, es decir, la capacidad de incorporar permanentemente nueva información y perfeccionar automáticamente sus recursos para analizarla, lo que permite a las máquinas generar su propio capital cognitivo.” (p.76)

Machine Learning. El Machine Learning es “la capacidad de un sistema de inteligencia artificial de adquirir su propio conocimiento a partir de patrones de datos sin procesar.” (Goodfellow, 2016, p.2)

Deep Learning. Uno de los tipos de Machine Learning más conocidos es el Deep Learning, que permite a una computadora “construir conceptos complejos a partir de conceptos simples.” (Goodfellow, 2016, p.2)

Generative Adversarial Networks. Los deepfakes se crean mediante un tipo de redes neuronales basadas en Deep Learning llamadas ‘*Generative Adversarial Networks*’ (GANs):

El modelo generativo es enfrentado contra un adversario: un modelo discriminador que aprende a determinar si una muestra es del modelo distribuidor o del distribuidor de datos. Este modelo generativo puede ser pensado como una analogía de un equipo de falsificadores, que trata de producir dinero falso y utilizarlo sin ser detectados, mientras un modelo discriminativo es análogo con la policía, que trata de detectar el dinero falsificado. La competencia en este juego hace que ambos equipos

mejoren sus métodos hasta que las falsificaciones son indistinguibles de los artículos genuinos. (Goodfellow, 2014, p.1)

Synthetic Media. El Synthetic Media está cambiando la forma en la que producimos contenido audiovisual. Esta tecnología permite que los usuarios creen contenido artificial que imita al creado por los seres humanos. Ajder y Glick (2021) definen el Synthetic Media como “contenido creado o modificado por una inteligencia artificial” (p.9)

Deepfakes. El termino deepfake es originado en 2017, por el nombre de un usuario de Reddit que desarrollo un método basado en el Deep Learning para sustituir una cara de un actor o actriz en videos pornográficos (Millière, 2022, p.1) Aunque el usuario de Reddit fue bloqueado, sus acciones desataron un creciente interés en los deepfakes, y nuevos contenidos se comenzaron a esparcir en otras plataformas como Twitter y 4chan. (Buo, 2020, p.1)

Gamage et al. (2022) se preguntaron como pensaban estas comunidades y descubrieron que entre las principales discusiones cambiaban a lo largo del tiempo:

En 2019 los usuarios discutían como las plataformas deberían actuar para mitigar la desinformación de los deepfakes. En 2020, los posts se trataban de discusiones sobre la creación de los deepfakes y sobre la detección de voces creadas con IA. En 2021, las publicaciones se centraban en pornografía y deepfakes de celebridades de un área de contribuidores: India.

Además, llegaron a la conclusión que durante el 2020 y 2021 los usuarios de Reddit usaban la plataforma como una test-bed para verificar la calidad de sus videos deepfakes. (p.8-9)

Amenazas de los deepfakes

La tecnología avanzó (y sigue avanzando) tan rápido que es muy simple y requiere menos esfuerzo crear *Synthetic Media*. Llegamos a un punto en donde al ojo humano le cuesta cada vez más distinguir entre lo real y lo sintético. En consecuencia de estos avances, afirma Zobaed et al.

(2021), “la creación de caras realistas de una persona que no existe o la alteración de un rostro es algo alcanzable.” (p.1)

Esto trae grandes problemas, el primero y más común: se puede crear videos de personas diciendo o haciendo cosas que nunca hicieron. Más allá de esto, el uso más común de los deepfakes para de Ruiter (2021) es en la pornografía, lo que “demuestra claramente el daño que esta tecnología puede hacer.” (p.1316)

Otros autores como Sam Gregory (2021) se preocupan más por las implicaciones de esta tecnología, principalmente en el “impacto en la veracidad y la confianza del material audiovisual” (p.6)

Estafas y Chantajes. Los estafadores podrían usar esta tecnología para robar una identidad de una persona para extraer recursos financieros. Por ejemplo, mediante la tecnología deepfake podemos “crear una personificación digital en tiempo real de un ejecutivo, que le puede pedir a un empleado que haga una transferencia urgente o que le provea información confidencial.” (Westerlund, p.43)

Estas estafas pueden afectar a empresas pero también a individuos. Por ejemplo, el ‘*cuento del tío*’ sería mucho más convincente: un estafador podría usar un deepfake para hacerse pasar por un familiar que necesita dinero con urgencia.

Erosión de la Verdad. El problema con las *fake news*, es que se distribuyen muy rápido en las redes sociales, pero cuando se desmienten, su refutación no tiene el mismo grado de difusión. El Council de Europa clasificó a la información engañosa en *misinformation*; *disinformation* y *malinformation*:

- Disinformation: información que es falsa y es creada deliberadamente para dañar a una persona, grupo social, organización o país.

- Misinformation: información que es falsa, pero no es creada con la intención de producir daño.
- Malinformation: información que está basada en la realidad, usada para generar daño en una persona, organización o país. (Wardle & Derakhshan, 2017, p.5)

Los deepfakes van a contribuir al aumento de la difusión de información falsa, que no solamente crea atmosferas tóxicas, sino que genera cada vez más desconfianza en las personas. Por ejemplo, una institución va a poder ser cuestionada en consecuencia de la difusión de un deepfake que comprometa su misión o estándares.

Pinhanez (2022), afirma que “las consecuencias de esta diseminación viral es el daño a compañías y productos, eventualmente consecuencias financieras, pero también es interferir con decisiones medicas como las campañas de vacunación, difundir curas falsas para enfermedades o provocar episodios de linchamiento”. (p.4)

Lo más preocupante, dice Ovadya (2019), es que “las personas crean cada vez menos en el contenido real. Si algún video puede ser resultado de manipulación, no hay nada que detenga a un político de deslegitimar un video legítimo que lo desacredite.”

A esto lo llamamos *Liar's Dividend*: el uso de la existencia de ‘Synthetic Media’ como una excusa para negar acusaciones de grabaciones o videos inapropiados afirmando que la fuente es falsa. El aumento del Liar's Dividend trae consigo el aumento de la desacreditación de opiniones o noticias e incluso puede provocar campañas de desinformación o fake news que sesguen el resultado de una determinada situación

Chesney & Citron (2019) afirman que al combinar el debilitamiento de la verdad con el debilitamiento de la confianza el autoritarismo va a ganar espacio. Así “los regímenes autoritarios

y los líderes con tendencias autoritarias se van a beneficiar cuando la verdad objetiva pierda su poder.” (p.1786)

De esta forma, este tipo de regímenes podrían utilizar la manipulación y las fake news para confundir a la población, manipular su pensamiento y alejarla de la verdad. Esto lleva a que no solo se ponga en peligro la verdad, sino también la confianza en la democracia.

Debilitamiento del Periodismo y los Medios de Comunicación. Cuando uno es consciente de que un video es deepfake, nuestra percepción no es igual que si no estuviéramos al tanto de esto. Es por eso por lo que los deepfakes van a contribuir también en la erosión de la fe en el contenido digital. La calidad, cantidad y la velocidad en la que se producen los deepfakes pueden generar una sobrecarga de información para los mecanismos fact-checking.

Esto va a tener un gran impacto tanto en el periodismo como en los medios de comunicación, porque la desinformación va a comenzar a ser moneda corriente. Es así, que “las organizaciones de noticias podrían dejar de reportar rápidamente algo como real por miedo a que la evidencia sobre estos termine por ser falsa.” (Chesney & Citron, 2019, p.1784)

De igual manera, esto también significa un desafío para los periodistas. La desinformación producida por los deepfakes, hace que una gran parte de su trabajo se base en determinar si un texto, audio y/o medio audiovisual es real o no.

Esto es importante, porque el emisor del mensaje va a ser muy importante en un futuro. Calvo & Aruguete (2020) comentan que “el modo en que valoramos la fuente que emite el mensaje afecta tanto nuestra interpretación del enunciado como la percepción de su veracidad.” (p.71)

En consecuencia, los medios de comunicación van a tener que invertir en tecnologías que reconozcan patrones y máquinas que detecten deepfakes. De esta forma, los medios van a garantizar una mayor veracidad y calidad de su información.

Erosión de las Instituciones. Los deepfakes pueden ser utilizados para manipular una elección, hacer que un político diga cosas que nunca dijo o como un arma de desinformación mediante propaganda.

Meneses (2021) afirma que “la desinformación creada mediante deepfakes va a ser un desafío durante las elecciones, teniendo en cuenta que para un ojo desentrenado un deepfake puede ser difícil de distinguir de un video real.” (p.2) Por ejemplo, Gideon Lichfield en la cuenta de Twitter del MIT Technology Review simuló una *entrevista* con Vladimir Putin en tiempo real.

El European Parliament et al. (2021) nos comentan sobre un futuro dominado por ‘*echo chambers*’, es decir “deepfakes difundidos mediante microtargeting, que van a tener un encuadre de las personas, que solamente van a creer en lo que ven en su visión del mundo” (p.33)

En este escenario, la confianza de la población en las instituciones políticas o en un gobierno se va a ver erosionada a nivel mundial. Además, generaría una mayor desestabilización en los países y el aumento de la división de la sociedad.

Hay autores como Ünver (2018) hablando sobre *Tecno-totalitarismos*, donde “la inteligencia artificial terminó silenciosamente con la democracia”. Además, da el ejemplo del gobierno chino, que “ya está usando estructuras algorítmicas para separar contenido crítico y contenido que tiene el potencial de generar movilización.” (p.9)

de Ruiter (2021) dice que “los deepfakes pueden ser utilizados para causar estragos en las Relaciones Internacionales y la política doméstica”. (p.1317) Con esto se refiere a que a través de un deepfake se puede presionar relaciones diplomáticas, incitar al conflicto armado e incluso producir desconfianza sobre la credibilidad de un metraje

Por ejemplo, un deepfake lanzado durante una reunión del Consejo de Seguridad de la ONU podría influenciar los votos en asuntos de seguridad internacional, o la manipulación de una entrevista con un líder extranjero podría generar una crisis diplomática.

También existe el peligro de que un deepfake cause una división en la opinión pública y un debilitamiento de la confianza de un gobierno. Los ciudadanos podrían llegar a creer que un líder está mintiendo, lo que llevaría a un descontento social y una disminución en los niveles de cooperación. Esto generaría que los líderes de otros países comiencen a desconfiar de la información que proviene de ese país.

Chesney & Citron (2019) por otro lado, hacen más énfasis en que “las agencias de inteligencia de los gobiernos más poderosos están en una buena posición de tomar decisiones inteligentes sobre el peso que les dan a potenciales fakes.” (p.1782)

En resumen, los deepfakes son una amenaza a las Relaciones Internacionales por la falta de credibilidad que estos pueden llegar a generar. Por lo tanto, es necesario que los gobiernos trabajen en forma conjunta para el desarrollo de herramientas de IA para detectar y eliminar deepfakes y así limitar el daño.

Los videos, voces y textos deepfakes podrían ser utilizados como evidencia real durante un procedimiento judicial. El European Parliament et al. (2021) sobre esto sostienen que “incluso cuando exista la validación de procedimientos de evidencia de audio y video para detectar deepfakes, la mera existencia de los deepfakes va a seguir influenciando testimonio”, porque las personas “van a seguir testificando en base a lo que vieron o escucharon de un deepfake fuera de la corte.” (p.32)

Además de la posible inclinación de un caso para un lado u otro por este tipo de manipulaciones, otra preocupación es que los deepfakes pueden generar desconfianza al no ser

reconocidos como evidencia fiable. Buo (2020) afirma que esto supondría “un mayor costo y tiempo para verificar y autenticar la evidencia antes que sea admisible en una corte”. (p.2)

En resumen, los deepfakes representan un riesgo para la erosión de la justicia, porque pueden ser usados para manipular decisiones judiciales, crear confusiones en la evidencia y provocar un mayor costo en los procedimientos legales.

Meneses (2021) afirma “En los viejos tiempos si querías amenazar a Estados Unidos debías tener 10 avionetas militares, armas nucleares y misiles de larga distancia. Hoy todo lo que necesitas es la habilidad de producir un video falso realista que pueda amenazar las elecciones.” (p.2)

Un estudio realizado por Tammekänd et al. (2020) dice que “los deepfakes pueden ser utilizados como mecanismo D&D por entidades militares durante tiempos de guerra y pueden desestabilizar la toma de decisiones.” (p.58) Es decir, utilizar los deepfakes como un arma de intimidación con la que una nación o grupo puede amenazar a otra con videos falsos para influenciar decisiones políticas o conflictos

Además, sabemos que las *fake news* no solamente intensifican los conflictos sociales, sino que también incitan a la violencia y a los actos terroristas. Por lo tanto, no sorprendería que se utilice la tecnología deepfake como un arma psicológica.

En síntesis, las consecuencias militares de los deepfakes son: caos, desinformación, desestabilización de las instituciones, pérdida de confianza de los líderes y daño a la población. Por lo tanto, los países deben estar preparados para detectar este tipo de videos y utilizar este tipo de tecnologías de forma responsable.

Beneficios de los deepfakes

A pesar de tener múltiples usos negativos, la tecnología deepfake tiene también su lado bueno. Los nuevos contenidos creados con Deep Learning requieren menos tiempo, esfuerzo, habilidades, conocimientos y recursos para ser producidos.

Entretenimiento. Una de las áreas que podría ser más beneficiada por la tecnología deepfake es sin dudas la del entretenimiento. Entre sus principales usos encontramos la mejora de las producciones cinematográficas, el desarrollo de videojuegos, la industria musical y la sátira.

Los deepfakes vienen para mejorar los efectos especiales, eso seguro. Sin embargo, también debemos tener en cuenta que van a bajar los costos de la producción de imágenes y videos. Esto representa un gran avance para la industria del cine, ya que le da la capacidad de realizar producciones audiovisuales sin la necesidad de tener un gran presupuesto.

Algunos autores como Westerlund (2019) afirman que, gracias a estas técnicas los “creadores de películas pueden recrear escenarios clásicos en películas, usar efectos especiales y técnicas de posproducción avanzada en edición de rostros y mejorar videos amateur a calidad profesional.”

Además, agrega que se “pueden crear voces digitales de actores que actores que la perdieron por una enfermedad o mejorar un metraje de una película en lugar de hacerlo de nuevo.” (p.41) Esto es posible gracias a la tecnología Text-To-Speech (TTS), haciendo que los productores puedan arreglar una línea errónea en los doblajes de voz con facilidad.

Los deepfakes también abren la puerta a la creación de escenas que reproduzcan viejas películas y la segmentación de escenarios realistas. Esto permite crear escenarios sin la necesidad de viajar a lugares físicos para filmar, y ahorrar así tiempo y recursos.

Otra cuestión a tener en cuenta, y tal vez la más controversial, es el uso de deepfakes para reemplazar actores. Si bien los efectos que se logran son increíbles, es importante aclarar que a

través de este tipo de aplicaciones se podrían explotar a actores, desplazarlos de trabajos y generar descontento en la industria.

Un ejemplo claro de la utilización de la tecnología deepfake en la industria del cine es en la serie “El Comandante Fort”, en la cual se entrenó un deepfake de Ricardo Fort utilizando aproximadamente 15.000 rostros de Ricardo e intercambiándolo con otro actor.

Otro ejemplo de la aplicación de una tecnología similar a la deepfake en la industria cinematográfica es el uso de CGI (Computer Generated Imagery) en ‘*Rápidos y Furiosos 7*’ para recrear escenas de *Paul Walker* en la película.

La tecnología deepfake en la industria de los videojuegos podría ser aplicada no solo para el desarrollo, sino también para mejorar la experiencia del jugador. No resulta difícil imaginar la creación de clones digitales del jugador mediante modelados 3D y experiencias multijugador con realidad virtual.

De igual manera, esto también trae aparejado algunos problemas. Por ejemplo, en 2021 los creadores del juego ‘The Witcher 3’ utilizaron técnicas de deepfake para utilizar las voces de los dobladores en un mod y no les pagaron alegando que esa no es su voz, es una modificación artificial.

En un futuro va a ser posible que tu artista favorita cante la canción que quieras, incluso si es de una persona que ya falleció. Millièrre (2022) dice que ya no se van a necesitar habilidades artísticas gracias a las técnicas de transferencia de estilo. Mediante estas técnicas se podría “hacer una canción de rock con la voz y el estilo de Elvis Presley usando un *variational autoencoder*” (p.9)

Esto va a cambiar completamente la industria de la música, permitiendo a los oyentes tener acceso a contenido exclusivo que sería casi imposible de conseguir normalmente. Se podría crear

nuevas versiones y colaboraciones entre artistas sin contar con la presencia física del artista, lo que lleva a una reducción de tiempo y recursos en las producciones.

Además, se puede llegar a cambiar la forma en que los artistas reciben ingresos de una canción. A través de diferentes programas que generan música de forma artificial se puede automatizar y repartir las tarifas por uso y participación de una canción. Esto significaría también que los artistas van a poder recibir ingresos por sus contribuciones sin tener que promocionar su trabajo.

La sátira, por lo general, es la forma más común en la que se difunden los deepfakes. La utilización de la tecnología deepfake para parodiar y criticar a figuras públicas son una vía barata y efectiva que utilizan los activistas para tratar de difundir un mensaje.

El auge de los deepfakes renovó la forma de hacer sátira en internet, lo cual impacta directamente en el activismo dado la facilidad para crearse y distribuirse. No obstante, hay que entender que este tipo de activismo lleva consigo un riesgo, por lo que hay que estar consciente de las posibles consecuencias antes de publicar un video satírico.

El European Parliament et al. (2021) advierten que “No todas las aplicaciones de los deepfakes en parodias y sátiras son inocentes: se ha demostrado que grupos extremistas pueden usar expresiones satíricas para encubrir el esparcimiento de su ideología.” (p.28)

Protección de la Identidad. El caso más famoso de la utilización de tecnología deepfake para la protección de la identidad es el del documental “Welcome To Chechenya”, en el cual personas que forman parte de la comunidad LGBT, y que sufrieron persecución en Rusia pudieron testificar su experiencia mediante el face-swapping con caras de voluntarios. Esto les permitió a las víctimas mostrar su testimonio sin que su identidad fuera revelada.

Investigaciones Médicas y Psicológicas. La innovación que trae consigo la Inteligencia Artificial puede cambiar las reglas del juego en el campo de la salud: nuevas formas de prevenir enfermedades, de tratar a las personas y reducir riesgos. Entre los ejemplos más fuertes podemos encontrar la creación de voz para mudos o sordos. de Ruiter (2021) incluso habla de “regenerar artificialmente la voz de personas que son incapaces de hablar debido a enfermedades como el ELA.” (p.1317)

Por otro lado, otros autores hablan de la investigación del uso de GANs para detectar anomalías en Rayos X y detectar señales de enfermedades (Buo, 2020, p.2) o de recrear digitalmente una extremidad apuntada (Westerlund, 2019, p.41)

La Organización Mundial de la Salud (OMS), desarrolló un modelo basado en Inteligencia Artificial que se llama *Florence*, en el que se busca ayudar a fumadores con su adicción al tabaco. Los usuarios hablan con Florence, que busca ayudarlos a dejar de fumar.

El European Parliament et al. (2021) comentan que ya existen “terapias experimentales para tratar desórdenes mentales como la ansiedad, mostrándole a los pacientes videos deepfakes de sí mismos donde parecen afrontar sus miedos y conseguir sus sueños,” (p.28)

Educación. Algunos hablan de la creación artificial de figuras históricas para propósitos educativos, para evitar lecturas aburridas y darle una nueva forma de enseñanza trayendo la historia a la vida real a la nueva generación de estudiantes. (Chesney y Citron, 2019, p.1769) Por ejemplo, en el Museo de Dalí en Florida, Estados Unidos se creó artificialmente a un Salvador Dalí que interactúa con sus visitantes en la exposición *‘Dalí lives’*.

Aunque no sean una herramienta deepfake, es importante destacar a los Large Language Models (LLM), como ChatGPT, que son modelos de IA capaces de mejorar también la educación. Por ejemplo, puede ayudar a los estudiantes en “el desarrollo de la lectura y escritura; [...] el

aprendizaje de lenguajes y tipos de escrituras; [...]la investigación y escritura de tareas; y en el desarrollo del pensamiento crítico” (Kascenci et al., 2023, p.2)

También puede ayudar a los profesores a crear “experiencias de aprendizaje personalizadas para sus estudiantes. Estos modelos analizan la escritura del estudiante y sus respuestas para proveer respuestas y sugerencias que se alineen a las necesidades del estudiante.” (Kascenci et al., 2023, p.3)

Otro caso de un modelo LLM es ‘*GALACTICA*’, un modelo entrenado con 48 millones de investigaciones científicas, que busca ayudar a los investigadores a explorar literatura, preguntar cuestiones científicas y escribir ciencia a través de un modelo que distingue entre investigaciones insignificantes e importantes.

Sin embargo, hay que tener cuidado con esto, porque se podría utilizar la tecnología deepfake para reescribir la historia y difamar a individuos que murieron y ya no pueden defenderse. Un ejemplo reciente es el caso Ruso, que está invirtiendo un esfuerzo significativo en reescribir su historia y su implicación en la Segunda Guerra mundial (Radchenko, 2020)

Usos Comerciales. Los deepfakes pueden funcionar para crear y optimizar métodos para la distribución de información a sus clientes y empleados. Entre sus aplicaciones más importantes podemos incluir la traducción automática de contenido.

Por ejemplo, en la campaña ‘*Malaria Must Die*’, David Beckham *habló* nueve lenguas en un video mediante la tecnología deepfake. Lo increíble de esto es que podemos hacer una campaña de publicidad completa basándonos en una sola grabación y llegar a mucho más público.

Los deepfakes también llegan a la industria de la moda. Marr (2022) habla de “modelos virtuales que muestran nuevos diseños de prendas en un desfile virtual, que cuesta una fracción de lo que costaría hacer un show real.”

Un ejemplo de estos modelos virtuales puede ser ‘*Lil Miquela*’, que en su página oficial la definen como una cantante de pop que tiene 19 años para siempre, activista social e icono de estilo, llamada por la revista times como la persona más influenciadora de internet. Esta modelo virtual ya desfiló para marcas como Prada, Calvin Klein, Burberry y Outdoor Voices.

Entre las empresas más importantes de generación de videos creados con Inteligencia Artificial encontramos a *Synthesia*. Esta página web ofrece videos de avatares realistas a los que podemos hacer decir lo que necesitemos.

El European Parliament et al. (2021) hablan de “mejorar experiencias digitales e interacciones entre humanos y computadoras”, donde “los usuarios podrían interactuar con una imitación de un conocido más fácilmente que con un extraño virtual.” (p.27) Vamos hacia un futuro donde los asistentes virtuales van a ser cada vez más realistas, que nos va a conocer mediante la recopilación de datos y nos va a dar información relevante para nosotros.

Deepfakes, Inteligencia Artificial y el Impacto en la Comunicación Política

Creación y Detección de los Deepfakes

El proceso de creación de un deepfake se basa en dos etapas: una primera etapa de detección, alineación y segmentación de una cara. Luego se pasa a la segunda etapa: el entrenamiento.

En esta fase dos redes neuronales (generador y discriminador) se enfrentan entre sí. El discriminador verifica la autenticidad de la imagen comparándola con otras con las que fue entrenada, determinando así si fue creada por el generador o es una imagen real. Tammekänd et al., 2020, p.47)

Hukkelås et al. (2019) proponen GANs con crecimiento progresivo: “una técnica de entrenamiento que genera rostros progresivamente, empezando por una resolución de 4x4 que se

incrementa hasta llegar a 1024x1024.” (p.3) A través de esta técnica podemos obtener caras cada vez más realistas y de mayor calidad.

Por lo general, es necesario grandes volúmenes de imágenes y videos para entrenar estos tipos de modelos que generan imágenes y videos realistas (Zobaed et al., 2021). Sin embargo, se espera que en un futuro los algoritmos GANs sean entrenados con bases de datos más pequeñas y sean capaces de producir deepfakes de mayor calidad. (Buo, 2020, p.2)

El problema de los deepfakes es que cada vez es más simple crearlos y cada vez es más difícil detectarlos. Si bien ya se comenzaron a desarrollar herramientas para detectar este tipo de metrajés, los falsificadores se adaptan rápidamente a los nuevos mecanismos. Estamos siendo testigo de una carrera armamentista entre los creadores de deepfakes y los mecanismos de detección. Estos mecanismos se dividen en manuales y automáticos.

Pinhanez et al. (2022) habla de la detección manual y la diferencia según imágenes, videos y audio:

- 1) Detección manual de las imágenes: se realiza buscando imperfecciones en las leyes de la física como sombras, o mediante la búsqueda de imágenes inversa para corroborar si una foto fue originada en un contexto diferente.
 - 2) Detección manual de videos: se utiliza las mismas técnicas que se utilizan en la detección de imágenes falsas. Para detectar videos falsos, se buscan irregularidades en los parpadeos o en objetos que aparecen y desaparecen.
 - 3) Detección manual basada en el audio: se verifica como suena la voz para el oído humano. Es decir, que el tiempo entre oraciones fluya normalmente, que el tono de la voz no cambie drásticamente, que la voz no suene robótica, y que no haya errores en la pronunciación.
- (p.1-2)

Por otro lado, otros autores creen que si bien en el corto plazo la detección manual puede funcionar, a gran escala la detección humana de videos falsos es inviable. Las máquinas van a perfeccionarse cada vez más y crear deepfakes cada vez más realistas. Es por esto por lo que es necesario crear sistemas que detecten inconsistencias comunes en los diferentes métodos de creación. Algunos de los métodos automáticos propuestos son los siguientes:

- 1) Reconocimiento del orador: reconocimiento basado en la identificación y verificación del orador con solo un audio como input.
- 2) Detección de voz viva: una tecnología capaz de detectar si el sonido de la voz viene de una persona viva o de un clip pregrabado.
- 3) Reconocimiento facial: el software utilizado para identificar personas en materiales fotográficos también puede ser aplicado a este tipo de metrajes.
- 4) Análisis facial: los investigadores están desarrollando algoritmos para prácticamente todos los puntos de referencia de la cara. Por ejemplo, la posición y movimiento de la nariz, boca u ojos.
- 5) Inconsistencias temporales: cuando una imagen está parcialmente modificada el algoritmo que crea el deepfake debe de alguna manera crear una transición entre el contenido original y la manipulación, esto suele resultar en un área borrosa. (European Parliament et al., 2021, p.17-19)
- 6) Métodos enfocados en la generación de artefactos: se centran en detectar artefactos de las tres etapas del proceso de generación de un deepfake. Busca inconsistencias en la alineación de la cara interior, la parte sintetizada en el algoritmo deepfake, y marcas en la cara exterior de un individuo. También usa diferentes posiciones de la cara para detectar puntos de referencias de la posición central de la cara y de la cara en general.

- 7) Señales psicológicas: mediante técnicas relacionados al *media forensics* se puede detectar señales psicológicas del comportamiento humano para detectar deepfakes.
- 8) Enfocados en Datos: este enfoque utiliza redes neuronales para detectar videos deepfake.
(Rowan & Pears, 2022, p.2)

La Inteligencia Artificial en la Comunicación Política

No existe mejor momento para ser político que la actualidad. A través de la tecnología es posible maximizar toda la efectividad de una campaña electoral: el Big Data, el Data Mining y la Inteligencia Artificial son las principales armas que están marcando la forma de hacer política en esta década. Por otro lado, autores como Girardi (2019) afirman que los tres elementos que modificarán la política en este siglo son “los sistemas cada vez más inteligentes, una tecnología integrada y una sociedad más cuantificada.” (p.78)

El filósofo Byung Chul-Han (2014) nos comenta que a partir del gran avance de la tecnología estamos en transición a la web 3.0, en la que se “hace posible un registro total de la vida. Ahora también nos vigilan las cosas que usamos diariamente”. Hoy en día “se registra cada clic que hacemos, cada palabra que introducimos en el buscador.” (Han, 2014, p.80-81)

Toda esta información es realmente valiosa, y un recurso que los países buscan explotar cada vez más. Nogués (2018) afirma que en la actualidad “la información es un bien como cualquier otro, algo que tiene valor y que cambia el mundo, y que se intercambia, se compra y se vende. Un bien que puede adulterarse.” (p.257)

Lo que nos debemos plantear entonces es la capacidad que tienen estas tecnologías de influir en nuestras vidas, y como poder utilizar sus beneficios sin que resulten en una amenaza para la sociedad. Akin Ünver (2018) afirma que vamos camino a una sociedad en la que “los humanos se convierten en una extensión de los commodities que compran” (p.6)

Autores como Girardi (2019), coinciden con Byung Chul-Han, y dicen que nos dirigimos a la época de la *psicopolítica digital*, que “avanza desde una vigilancia pasiva hacia un control activo”, y que mediante el Big Data se logra un conocimiento “que permite intervenir en la psique y condicionarla a un nivel pre-reflexivo.” (p.80)

Uno de los mayores desafíos al que las sociedades se enfrentan hoy es a la desinformación. Pinhanez et al. (2022) la definen como “información falsa que se difunde con una intención de dañar.” (p.6)

En el lapso 2017-2019 se dio, según Tammekänd et al. (2020) un “150% de aumento en campañas de desinformación” y se estima que alrededor de “8 millones de cuentas en Twitter compartían desinformación por semana” (p.6) Estas campañas afectan directamente a la gente de mayor edad, según Barari (2021), “los usuarios de más de 65 años compartieron 7 veces más artículos de fake news que gente más joven durante la elección presidencial de Estados Unidos en 2016”. (p.4)

Los deepfakes ya están dando de qué hablar y se están usando para desacreditar evidencia verídica: aunque exista un video de alguien haciendo o diciendo algo, se puede afirmar que es un deepfake y es muy difícil demostrar lo contrario. Nogués (2018) hablando de la posverdad afirma que “la posverdad genera posverdad”, y que “si no podemos confiar en nada ni en nadie, seguimos lo que nuestra intuición nos dice, sumidos en todos nuestros sesgos y distorsiones.” (p.299)

Además de lograr que se empiece a jugar con el Liar’s Dividend y que se ponga en jaque la veracidad de los contenidos con facilidad a través de los deepfakes, algunos “líderes autoritarios y los movimientos que los apoyan pueden utilizar el humor como una espada y escudo, para atacar a sus enemigos, mientras que se defienden afirmando que «es una broma».” (Adjer y Glick, 2021, p.7)

Por ejemplo, en dos incidentes políticos recientes en Malasia y Brasil, “se buscó evadir las consecuencias de un video comprometedor afirmando que esos videos eran deepfake. En ambos casos, nadie fue capaz de establecer lo contrario, y la opinión pública permaneció dividida.” (Toews, 2020)

Durante la historia reciente, la política tuvo una serie de herramientas que le ayudaron a ver el estado general de la sociedad. La utilización del Big Data por parte de empresas como Cambridge Analytica es una de las demostraciones más claras de la capacidad de manipulación que se puede lograr a través de los datos.

No es ninguna novedad el uso de microtargeting en campañas políticas, es decir, “dirigirse con precisión a los electores, con mensajes personalizados para influirlos” (Han, 2014, p.81). Esto fue y es un problema, pues “cuando salen a luz los intentos de manipular la opinión pública por parte de algunos grupos de poder, uno de los daños colaterales es que todos los políticos pierden credibilidad ante nosotros.” (Nogués, 2018, p.299)

Los sistemas gubernamentales ya intentaron localizar campañas automatizadas de influencia anteriormente. En 2017, “investigadores descubrieron que un millón de comentarios subidos a la ‘Federal Communications Commission’ eran generados automáticamente, con ciertas frases copiadas y pegadas en diferentes mensajes” (Knight, 2019)

Si a estas estrategias de microtargeting, le agregamos la automatización, tenemos la fórmula perfecta para un futuro distópico. A diferencia de las fake news que conocemos hoy, en un futuro será necesario que la prevención de fake news se realice de forma automática.

Los campos que más se ven afectados son la comunicación en general: la propaganda generada con IA, que puede ser “desarrollada a escala, mientras es personalizada y muy poco

costosa de producir. Además, puede ser distribuida ampliamente a través de plataformas como redes sociales para conducir a operaciones de información.” (Pinhanez et al, 2022, p.22)

Esto va a lograr que las personas se queden cada vez más en sus burbujas políticas, ya que van a recibir contenido creado exclusivamente para ellos a través de una Inteligencia Artificial, que va a estar alineado a sus creencias y va a confirmar sus pensamientos.

Hoy, la mayoría de las Fake News son escritas por humanos, pero indicadores tempranos apuntan a que en un futuro “la desinformación va a ser creada a un costo casi nulo y con una velocidad exponencial” (p.22) a través de los LLM (Large Language Models) como GALACTICA o GPT-4, que han alcanzado a un punto en que a través de una petición (prompt) “pueden escribir artículos que parecen al menos superficialmente convincentes.” (Ovadya y Whittlestone, 2019, p.2)

Sin embargo, la IA también puede ser utilizada para que las campañas tengan mayor legitimidad. En un enfoque ético, la IA puede “trabajar para informar y servir al electorado. Nuevas start-ups de IA como Factmata y Avantgarde Analytics ya están proveyendo estas soluciones tecnológicas.” Por ejemplo, es posible programar bots políticos para que las personas dejen de compartir información falsa. Esto se podría lograr a través de advertencias de que la información es sospechosa y por qué.” (Polonski, 2017)

La automatización lleva a pensar que en algún momento se llegará a un presidente que se base en decisiones tomadas por una Inteligencia Artificial. Sin embargo, Akin Ünver (2018) comenta que en este pensamiento se “omite tal vez la fuente más importante de autoridad en la política: la legitimidad.” (Akin Ünver, 2018, p.8)

Sobre esto, profundiza incluso más y se pregunta: “Cuando un algoritmo tome una decisión automática de ir a la guerra, o de incrementar el presupuesto militar, ¿Cómo va a

conseguir legitimidad para esta decisión?.” (Akin Ünver, 2018, p.8) Tal vez el mejor uso de la IA en los espacios políticos será la “capacidad de mejorar la eficiencia de los costos y proveer lo mejores estimativos de costos de materiales y trabajos para considerar otras alternativas.” (Akin Ünver, 2018, p.10)

Carrera Armamentística de la Inteligencia Artificial

La carrera armamentística de la IA ya comenzó, incluso algunos la consideran una nueva guerra fría. Presidentes de grandes potencias ya manifestaron su interés en el desarrollo de la Inteligencia Artificial en sus países: los casos más destacados son el de Vladimir Putin en Rusia y Emmanuel Macron en Francia.

Rusia. El presidente ruso afirmó que “el país que desarrolle mejor su IA en el corto plazo va a ser el que controle el mundo.” Elon Musk, sobre esto, opinó en Twitter que “la competición por la superioridad en IA a niveles nacionales va a causar la Tercera Guerra Mundial.” (Akin Ünver, 2018, p.5)

No es la primera vez que Vladimir Putin habla sobre la importancia de la IA, incluso en ese mismo discurso afirmó que “a Inteligencia Artificial es el futuro, no solo para Rusia, sino para toda la raza humana”, y que “trae consigo oportunidades colosales, pero también amenazas que la hacen difícil de predecir.” (Ratner, 2017)

Francia. Tanto Vladimir Putin como Elon Musk se desvían de la posición de Emmanuel Macron, que se enfoca principalmente en “el sistema de salud, las finanzas y la participación política en aspectos de estructuras algorítmicas, más que en aplicaciones militares.” (Akin Ünver, 2018, p.5)

En una entrevista para WIRED, Emmanuel Macron declaró que el gobierno francés va a invertir “1.5 billones de euros en cinco años para apoyar a la investigación en el campo, alentar a

las startups, y recolectar información que pueda ser usada, y compartida por ingenieros.” (Thompson, 2018)

Además, en esta misma entrevista el presidente francés declara que la IA va a “ser disruptiva en todos los tipos de negocios” y que Francia debe estar preparada. Por esta razón, su principal objetivo en “términos de educación, entrenamiento, investigación y creación de startups es tener sistemas adaptados, finanzas adaptadas y regulaciones adaptadas, para así poder ser los mejores en el campo y atraer a los mejores”. (Thompson, 2018)

En relación a las demás potencias, Macron aclara que “Europa no tiene las mismas preferencias colectivas que Estados Unidos o China” (Thompson, 2018), y que si quieren defender su forma de enfrentar la tecnología deben ser una parte activa de la revolución en Inteligencia Artificial.

Los Deepfakes en la Comunicación Política

Según Ovadya y Whittlestone (2019) “un ejemplo específico en el que el mal uso se ha incrementado es el uso del Machine Learning en el Synthetic Media: la manipulación y generación de audios, videos, imágenes y texto realista.” (p.2)

El uso de los deepfakes ha dado lugar a situaciones controvertidas en el campo de la política, como metrajes falsos de funcionarios diciendo cosas que nunca dijeron o videos de políticos haciendo campañas en diferentes idiomas. Los deepfakes comenzaron a tener implicaciones en la comunicación política, y los políticos ahora deben empezar a preocuparse por los nuevos desafíos que trae consigo esta tecnología.

Hoy en día es posible producir imágenes de caras que son casi indistinguibles de fotografías. “Se está volviendo cada vez más fácil manipular videos existentes de personas, reemplazar las expresiones faciales y movimiento con los de otras personas.” (Ovadya y

Whittlestone, 2019, p.2) A través de softwares como DeepFaceLab y Faceswap “cualquier con tiempo y acceso al cloud computing puede desarrollar procesos de Machine Learning y renderización grafica sin un conocimiento previo.” (Puutio, 2020)

Como se puede esperar, el número de videos deepfakes está “creciendo exponencialmente mientras continua la democratización de la IA y el cloud computing logra que los procesos sean cada vez más accesibles.” (Puutio, 2020) El estudio de Tammekänd et al. (2020) afirma que hay “más de 100 millones de deepfakes online” y un “aumento de x6820 desde 2019”. (p.7)

Con anterioridad, se mencionó que “los deepfakes pueden ser abusados para crear tensiones éticas o políticas entre países, para intentar engañar al común de las personas para alterar una elección, o para crear el caos en deportes o la economía global creando contenidos falsos.” (Zobaed et al., 2021, p.2). Sin embargo, también debemos tener en cuenta que lo realmente importante es la tecnología que está detrás de los deepfakes. Este tipos de tecnología sin dudas va a fortalecer la comunicación inmersiva. Mosley (2019) dice que “si alguna vez quisimos tener una comunicación inmersiva, debemos tener la capacidad de crear apariencias foto-realistas de nosotros mismos en orden de lograrlo.”

Cambios en las Formas de Comunicación

La forma que interpretamos un texto se ve afectada por la valoración que hacemos de la fuente, Calvo y Aruguete (2020), dicen que “el modo en que valoramos la fuente que emite el mensaje afecta tanto nuestra interpretación del enunciado como la percepción de su veracidad” (p.71). En estos tiempos donde los deepfakes y sus posibles implicaciones en la sociedad generan tantas dudas, existe la posibilidad de que esto sea una oportunidad para lograr “la restauración de la confianza en los periodistas, como unos guardianes de la sociopolítica en la era digital de la posverdad.” (Gregory, 2021, p.8)

En un estudio realizado por Barari (2021), se llegó a la conclusión que aunque los deepfakes “sean diseñados específicamente para difamar a un político, no son únicamente creíbles o emocionalmente manipulativos: no son más efectivos para desinformar como la presentada en texto o audio a los mismos objetivos atacados vía una campaña de anuncios”. (Barari, 2021, p.12)

La posterior investigación de Groh et al. (2022) afirma que si bien “las investigaciones demuestran que los videos, especialmente de injusticias, tienen más llegada y reacciones emocionales que descripciones de textos mostrando la misma información” (p.1), sus resultados “contrastan con la confianza convencional de que los videos son más persuasivos que el texto para convencer a la audiencia de un evento falso. Encontramos que los participantes evaluaron mejor la autenticidad de un discursos políticos visuales que textuales.” (p.7)

Aunque estas investigaciones sean recientes, es innegable el avance que están logrando los deepfakes. Por lo que cuando se vuelvan indistinguibles, seguramente habrá una mayor dificultad para verificar la autenticidad de los mensajes visuales que se reciben y se comparten en línea.

La propagación rápida de las fake news es una de las razones por las cuales la posverdad se ha vuelto tan efectiva en nuestra sociedad. A menudo la versión real de los hechos no consigue la misma difusión que las noticias falsas, lo que convierte a la verdad en una víctima de la propaganda.

Como respuesta a este problema, se han propuesto soluciones como la automatización del fact-checking. Sin embargo, es importante tener en cuenta que los algoritmos tienen limitaciones y pueden generar errores. A pesar de que la inteligencia artificial se basa en datos, puede interpretarlos de manera equivocada, lo que significa que la revisión humana siempre será necesaria.

En este sentido, la autenticación de la humanidad y la identidad virtual adquieren una importancia cada vez mayor. La tecnología ha evolucionado rápidamente y, con ella, también los métodos para engañar a los usuarios. Por eso, la verificación de identidad se ha vuelto crucial para garantizar que la información no sea escrita por bots. Hoy en día, nuestra identidad en línea es tan valiosa como nuestra identidad física, por lo que debemos protegerla con todo el cuidado que merece.

Casos de Deepfakes en la Política

En esta investigación se viene siguiendo una línea similar a lo planteado por Ai et al. (2022): “la información falsa que confunde al público no solo intensifica conflictos sociales e incita a la violencia y acciones terroristas, sino que también se convierte en un arma psicológica en el campo de batalla.” (p.1)

Los deepfakes ya son utilizados en la política actual. Por ejemplo, en Gabón, el presidente Ali Bongo Ondimba tenía mucho tiempo sin una aparición pública, por lo que el gobierno decidió lanzar un video para “calmar los rumores de que estaba enfermo o había muerto. Se sospechaba que este video era un deepfake por los movimientos extraños en los ojos del presidente.” (Tammekänd et al., 2020, p.24) Este video resultó en un intento de golpe militar: “Los militares pensaron que el video era falso, sin embargo el presidente confirmó luego que era real” (Meneses, 2021, p.5)

En India, un día antes de las elecciones legislativas de Delhi, dos videos de Manoj Tiwari pidiendo a los ciudadanos que voten por su partido aparecieron en internet. (Meneses, 2021, p.5) En estos videos el presidente hablaba diferentes lenguas para llegar a más cantidad de votantes. Según Tammekänd et al. (2020) el video llegó a “15 millones de personas y 5800 grupos de WhatsApp”. (p.25)

En 2021, luego del golpe militar, se lanzó un video en Myanmar en el que un prisionero político acusaba al presidente de facto, Aung San Suu Kyi de corrupción. En el video, Phyo Min Thein, está sentado en una habitación leyendo una declaración. Su habla parece extraña y no es su voz normal, su cara esta estática y sus labios parecen fuera de sincronización. (Gregory, 2021)

Gregory (2021) asegura que “el público y la prensa en Myanmar alegaban que era un deepfake, basándose principalmente en evaluaciones visuales y en resultados de detectores online poco fiables.” Aunque probablemente el video no era un deepfake y era más bien una declaración forzada, “la sociedad civil y los periodistas no fueron capaces de acceder a las habilidades y herramientas necesarias para verificarlo.” (p.9)

Otro ejemplo puede ser el deepfake de Xi Jinping como Winnie the Pooh, en el que se muestra al presidente chino “usando un disfraz del oso de la ficción, una comparación que se volvió un símbolo de resistencia contra el gobierno comunista de Xi” (Ajder y Glick, 2021, p.15) o el caso del deepfake de Donald Trump lanzado por un partido político belga en el que el expresidente de Estados Unidos pide que Bélgica abandone el Acuerdo de Paris (Tammekänd et al., 2020, p.24)

Regulaciones y Soluciones

Regulaciones

El problema que existe a día de hoy con las regulaciones es la no contemplación del ‘Synthetic Media’, por lo que se puede alegar que este tipo de manipulaciones no representan a una persona sino que son ‘creaciones digitales’.

Indudablemente, los gobiernos deberían comenzar a desarrollar rápidamente políticas para promover el uso apropiado de la inteligencia artificial en la creación de contenido y el apoyo al desarrollo tecnológico para verificar la autenticidad del video y el contenido de audio.

Chesney & Citron (2019) dicen sobre esto que “crear una ley que prohíba la utilización destructiva del Synthetic Media tratando de excluir sus usos beneficiosos sería difícil, pero no imposible” (p. 1789)

Algunos países ya vienen pensando en regular esta tecnología, las legislaciones más importantes son las de Europa, Estados Unidos (Texas y California) y China. Además, al entender al internet como un lugar donde no hay leyes que regulen lo que sucede, es importante también que las plataformas apliquen políticas contra esta tecnología.

Europa

El European Parliament Research Service realizó una importante investigación sobre los deepfakes para la creación de un marco regulador que se basa en el riesgo para regular las Inteligencias Artificiales.

“La comisión distingue entre ‘riesgo inaceptable’, ‘alto riesgo’, ‘riesgo limitado’, y ‘riesgo mínimo’. Con esta propuesta, la comisión busca prohibir el uso de sistemas de IA que lleven consigo un riesgo inaceptable hacia la seguridad y los derechos fundamentales de los ciudadanos europeos.” (European Parliament et al., 2021, p.38)

Algunas opciones que se plantearon son:

- Regular la tecnología deepfake como de alto riesgo: “regular la tecnología deepfake como un sistema de IA de alto riesgo bajo la estructura propuesta para IA” (European Parliament et al., 2021, p.59)
- Creación de obligaciones legales a los proveedores de deepfakes: hoy en día no hay obligaciones para que los proveedores tengan que etiquetar un contenido deepfake. Una política podría ser “extender la estructura de la propuesta en IA para obligar a los

productores de la tecnología incorporar etiquetados.” (European Parliament et al., 2021, p.59)

- Prohibir los deepfakes en propaganda política o comunicación: las investigaciones indican que “los deepfakes combinados con microtargeting podrían ser utilizados para manipular las opiniones políticas.” (European Parliament et al., 2021, p.61)

Estados Unidos

En Estados Unidos se comenzó a legislar en algunos estados. Las legislaciones más importantes son las de California y Texas

California (AB 730). Este proyecto de ley busca “prohibir a una persona, comité u otra entidad, distribuir contenido malicioso (audio o videos) que intente perjudicar a la reputación del candidato que va a aparecer en una boleta entre los 60 días de una elección, a menos que el contenido incluya un aviso de que el contenido fue manipulado”

Texas (S.B. NO. 751). El proyecto de ley provee excepciones a medios de comunicación y a contenido que pueda ser considerado sátira o parodia.

“Una persona que comete una ofensa a una persona con la intención de dañar a un candidato o influenciar el resultado de una elección:

- (1) Crea un video deepfake; y
- (2) Causa que el video deepfake sea publicado o distribuido entre los 30 días de una elección”

China

El caso chino es el más importante, teniendo en cuenta que es un país que siempre está a la vanguardia de las legislaciones tecnológicas. En el caso de los deepfakes la CAC (Cyberspace Administration) publicó regulaciones que “prohibían la creación sin el permiso del sujeto o que sea considerado en contra de los intereses nacional. Cualquier cosa que vaya en contra de los

valores socialistas recae en esta descripción, como cualquier forma de información ilegal o maliciosa” (Sharwood, 2022)

Además, ahora “las identidades online tienen que ser autenticadas por las plataformas, y los usuarios pueden reportar contenido falso sospechoso” (Quach, 2023)

Redes Sociales

Es indispensable que el Estado construya confianza mutua entre investigadores, privados, start-ups y ciudadanos. Los creadores de políticas deben trabajar en conjunto con las compañías de redes sociales para desarrollar estándares de contenido.

Hoy en día, algunas plataformas ya comenzaron a tomar medidas:

- Facebook: anunció que “comenzaría a remover contenido que fue editado o sintetizado de maneras que no son aparentes en una persona promedio y que podría llevar a pensar a alguien que un sujeto en un video dijo algo que en realidad no dijo, y que fueran generado usando IA o ML.” (Hwang, 2020, p.14)
- Twitter: anunció que “removería y advertiría a los usuarios contra el contenido manipulado o sintético que sean probables de causar daño” (Hwang, 2020, p.14)
- Reddit: además de eliminar el subreddit r/deepfakes, viene adoptando políticas similares a Facebook y Twitter.
- TikTok: sus lineamientos prohíben “falsificaciones digitales (Synthetic Media o contenido manipulado) que lleven a los usuarios a distorsionar la verdad de eventos y que causen daño al sujeto del video, otras personas o a la sociedad”. (Ajder, 2021, p.43)
- YouTube: sus políticas prohíben “el contenido manipulado, que la plataforma lo define como contenido que fue técnicamente manipulado de manera que confunda a los usuarios y que pueda generar un riesgo serio de un enorme daño” (Ajder, 2021, p.43-44)

Soluciones

Más allá de las soluciones legales, algunos autores proponen soluciones alternativas al problema de los deepfakes. Estas incluyen la utilización de inteligencia artificial para detectar los deepfakes, la verificación de fuentes de contenido para ayudar a determinar si un video es real o un deepfake, y la educación para ayudar a la gente a reconocer cuándo se encuentran frente a deepfakes. A continuación, se explican algunas de estas soluciones alternativas:

Soluciones Mediante Autenticación

Esta solución permite a los usuarios verificar la validez de un contenido mediante publicaciones o capturas verificadas. Hace tres años, Adobe lanzó ‘Content Authenticity Initiative’, una iniciativa que detalla “como de original es el contenido, manipulaciones, cambios, ediciones y agregados que pueden ser rastreados en fotos y videos.” (Gregory, 2021, p.10)

Lapowsy (2019) comenta sobre NewsGuard, la extensión que “fija un cono rojo a los sitios web que fallan su test y un icono verde a los que la pasan. Los usuarios pueden ver una etiqueta general que explica cada ranking.”

Sobre estos fact checking sites, Rodríguez-Ferrandiz (2019) comenta que “mientras algunos sostienen que en ciertos casos refuerzan más que corrigen aquello cuya veracidad discuten, otros destacan su capacidad para revertir la situación y servir a los fines de una pedagogía informativa.” (p.7)

Sin embargo, nada garantiza que los sitios de fact-check tengan una línea editorial imparcial, por lo que los usuarios no deberían depender exclusivamente de ellos. La verificación cruzada, revisando diferentes fuentes, resulta el medio más seguro entonces para determinar la veracidad de la información.

Soluciones con Tecnología

Pinhanez et. al (2022) afirman que una solución podría ser un protector de contenido falso: “un software instalado en los dispositivos de los usuarios que pueda verificar el contenido al que el usuario está expuesto y que advierta de los posibles problemas con el contenido” (p.2)

Mediante este protector, los usuarios podrían reportar contenido fraudulento, bloquear contenido sospechoso y marcar contenido verificado. Esta solución podría ayudar a los usuarios para discernir entre contenido real y falso.

Hukkelås et al. (2019) proponen un modelo de GAN condicional que “genera imágenes basadas en el contorno de la cara y que separa la información de la posición.” (p.5)

El problema de estos softwares es que, como dicen Chesney & Citron (2019), “los mecanismos de detección deberían seguir el avance de las innovaciones de los deepfakes para retener su eficacia.” Y aún si existiese este tipo de tecnología y pudiese ser desarrollada en las redes sociales, “solamente reduciría el daño, pero no lo eliminaría.” (p.1787)

Es por esto por lo que, aunque la tecnología pueda ayudar a prevenir el contenido falso, los usuarios también deben comprender los peligros de internet y aprender a identificar el contenido manipulado.

Soluciones con la Blockchain

Hasan & Salah (2019) proponen una solución basada en el rastreo y transparencia antes que en la detección. Mediante el uso de la tecnología blockchain proponen que “El dueño de un video cree un contrato inteligente donde otros artistas pueden pedir permiso para editar, alterar o distribuir siguiendo los términos y condiciones de un acuerdo.” (p. 41600)

Fraga-Lamas & Fernandez-Caramés (2019) dicen que “DLTs como Tangle o la blockchain son capaces de proveer autenticaciones eficientes y almacenamiento de datos seguros, procesando y compartiendo, escalabilidad, transparencia y responsabilidad.” (p.3)

Otras Soluciones

Galston (2020) propone la formación de un “equipo anti-deepfake SWAT no partidario que reúna a los mejores académicos e investigadores del sector privado, con una línea manejada 24/7.”

Ovadya (2019) dice que la mejor defensa contra los deepfakes es también la mejor defensa contra los cheapfakes, “asegurarse que plataformas como Facebook y YouTube eviten recompensar cualquier forma de información falsa o maliciosa con atención o redito.”

Ética

Un deepfake puede ser usado para el bien o para el mal. Las tecnologías en esto son neutrales y las implicancias políticas que estas generan son producto de la personas. El uso ético de los deepfakes entonces, va a depender de la forma en la que es utilizado.

Teniendo en cuenta esto, es importante que los deepfakes sean utilizados de la manera más responsable posible. Por lo tanto, el uso ético de un deepfake depende en gran medida de la intención de quien lo está utilizando y de los efectos que tenga sobre otros.

Sobre esto, el creador de la cuenta *DeepfakesAR* comenta que los deepfakes no son éticos cuando “se usan para engañar o dañar a otros, ya sea de manera intencional o no. Por ejemplo, si se usan para difundir noticias falsas o manipular videos con fines políticos o financieros.” Y agrega, que “si se usan para fines educativos o artísticos, puede ser considerado aceptable. Pero, es importante tener en cuenta que si no se tiene el consentimiento de las personas retratadas, puede violar sus derechos personales y privacidad.”

Según De Ruiter (2021), hay tres factores centrales a la hora de determinar si un deepfake es moralmente problemático: “(i) si la persona a la que se le hizo el deepfake está en desacuerdo de la forma en la que se lo representa; (ii) si el deepfake engaña a los que lo ven; y (iii) la intención con la que el deepfake fue creado.” (p.1)

De Ruiter (2021), dice que “los deepfakes pueden ser moralmente aceptables si están realizados con el consentimiento de la persona representada” (p.12)

Ante la posibilidad de que la gente se vea engañada, algunas plataformas como Deep Reckonings están buscando hacer explícitos que los videos son Fake mediante los créditos, los guiones y las marcas de agua.

Por otro lado, algunos autores justifican el uso de los deepfakes en la sátira. Esta justificación se da en un contexto de gran diferencia de poder: la sátira se convierte en una ‘*weapon of the weak*’: una herramienta para alzar la voz de los débiles en espacios donde no son tomados en serio. De este modo, los deepfakes puede ser utilizados como forma de expresión para personas sin acceso a medios de comunicación más grandes. Por ejemplo, muchos falsifican videos para abordar temas ignorados por los medios.

Es importante recordar que los deepfakes son simplemente una evolución de las tecnologías de edición de video, donde muchas veces, sin darnos cuenta, nos exponemos a contenidos falsos o manipulados sin saberlo. Por lo tanto, es importante también ser conscientes y responsables a la hora de consumir y difundir cualquier contenido digital.

Conclusión

Los deepfakes llegaron para quedarse, y con el tiempo se nos va a hacer cada vez más difícil distinguir lo real de lo sintético. Estamos llegando a un punto en la ciberseguridad se está convirtiendo en un ‘*cat and mouse game*’, con constantes mejoras en la seguridad para combatir los constantes ciberataques. Las tendencias de desarrollo en la actualidad es generar *outputs* de alta calidad basados en *inputs* muy pequeños.

Vamos hacia un futuro en el cual las nuevas tecnologías nos van a permitir usar la nube para manipular videos en streaming, donde las compañías provean los deepfakes como un producto

o un servicio online. A mi parecer, en un futuro no muy lejano, será necesario dejar por sentado cuando un producto o servicio fue realizado por una Inteligencia Artificial o no.

Tenemos que estar preparados para un futuro en el cual los deepfakes sean moneda corriente, por lo tanto es realmente importante que se fomente el desarrollo y utilización de herramientas y tecnologías que sirvan para identificar y combatir los deepfakes.

Los asistentes virtuales van a desarrollarse a tal punto que van a ser una extensión de nuestra vida. Vamos a poder interactuar con un asistente similar a un humano, que nos va a permitir interactuar y comunicarnos con este, logrando que nos olvidemos de que en realidad, es un robot. Muchas cosas que hoy consideramos ciencia ficción, dentro de un futuro no muy lejano se van a convertir en realidad.

Cuando pensamos en las cuestiones éticas debemos tener en claro que un deepfake puede ser utilizado tanto para el bien como para el mal. Sin embargo, no debemos olvidar que la tecnología en esto es neutral, y las implicaciones políticas que estas generan son productos de las personas. Por lo tanto, la mejor arma para prevenir el impacto del uso malintencionado de los deepfakes es educar a la gente para que pueda reconocer este tipo de contenido manipulado.

Esta tecnología ya está en el mundo digital, y se desarrolla cada vez más rápido. Es necesario crear un ambiente de investigación más seguro y mitigar los daños que producen los deepfakes.

Los avances en la tecnología deepfake va a permitir que seamos capaces de crear y consumir contenido de alta calidad, que hoy en día resulta imposible de hacer. Al mismo tiempo, nos va a permitir expresar con mayor claridad nuestras ideas.

Indudablemente, la tecnología y los avances digitales van a seguir siendo parte de nuestras vidas. Estamos a las puertas de un futuro en el cual vamos a ser capaces de llevar cualquier cosa a un

nivel superior gracias a la tecnología. La visión que debemos tener a futuro no debería ser solo la de conseguir el progreso tecnológico, sino el progreso de la especie humana.

Referencias Bibliográficas

- Ai, J., Wang, Z., IEEE, Huang, B., Han, Z. (2022). *Deepfake Traceability with Disentangling Reversing Network*. ArXiv. <https://doi.org/10.48550/arXiv.2207.03666>
- Ajder, H. y Glick, J. (2021). *Just Joking! Deepfakes, satire, and the politics of synthetic media*. WITNESS & MIT Open Documentary Lab. <https://cocreationstudio.mit.edu/wp-content/uploads/2021/12/JustJoking.pdf>
- Akin Ünver, H. (2018). *Artificial Intelligence, Authoritarianism, and the Future of Political Systems*. Centre for Economics and Foreign Policy Studies. <http://www.jstor.org/stable/resrep26084>
- Barari, S., Lucas, C. y Munger, K. (2021). *Political Deepfakes Are as Credible as Other Fake Media And (Sometimes) Real Media*. <https://doi.org/10.31219/osf.io/cdfh3>
- Bellomo, S. y Oszlak, O. (2020). *Desafíos de la Administración Pública en el Contexto de la Revolución 4.0*. Konrad Adenauer Stiftung. https://www.kas.de/documents/287460/4262432/desafios_revolucion_4_0.pdf
- Buo, A. S. (2020). *The Emerging threats of Deepfake Attacks and Countermeasures*. ArXiv. <https://doi.org/10.48550/arXiv.2012.07989>
- Calvo, E. y Aruguete, N. (2020). *Fake News, trolls y otros encantos*. Siglo XXI
- Chesney, B., y Citron, D. (2019). *Deep fakes: A looming challenge for privacy, democracy, and national security*. California Law Review, 107(6), 1753–1820. <https://doi.org/10.15779/Z38RV0D15J>
- de Ruiter, A. *The Distinct Wrong of Deepfakes*. Philos. Technol. 34, 1311–1332 (2021). <https://doi.org/10.1007/s13347-021-00459-2>

- European Parliament, Directorate-General for Parliamentary Research Services, Huijstee, M., Boheemen, P., Das, D. (2022). *Tackling deepfakes in European policy*. European Parliament. <https://data.europa.eu/doi/10.2861/325063>
- Fraga-Lamas, P. & Fernández-Caramés, T. M. (2019). *Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality*. Universidade da Coruña. <https://doi.org/10.48550/arXiv.1904.05386>
- Galston, W. A. (2020). *Is seeing still believing? The deepfake challenge to truth in politics*. Brookings. <https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/>
- Gamage, D., Ghasiya, P., Bonagiri, V. K., Whiting, M. E. y Sasahara, K. (2022). *Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications*. ArXiv. <https://doi.org/10.48550/arXiv.2203.15044>
- Girardi, E. (2019). Digitalización, política e inteligencia artificial ¿Qué futuro podemos esperar? *Revista Nueva Sociedad*, Volumen 283, 75-81. https://static.nuso.org/media/articles/downloads/6.TC_Girardi_283.pdf
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. y Bengio, Y. (2014). *Generative Adversarial Nets*. Université de Montréal. <https://doi.org/10.48550/arXiv.1406.2661>
- Goodfellow, I., Bengio, Y. y Courville, A. (2016). *Deep Learning*. MIT. <https://www.deeplearningbook.org>

- Gregory, S., Posters, B., Howe, D., Lepp, S., Cizek, K. [WITNESS] (2020, Agosto 28) ‘*Faking the Powerful*’ with deepfakes: with Bill Posters, Daniel Howe, and Stephanie Lepp. [Video]. https://youtu.be/yff_IAWW57E
- Gregory, S. [Truth and Trust Online] (2020, Noviembre 17). [Part 2] *Manipulated media detection and mitigation, what to prioritize and who to prioritize* [Video]. <https://youtu.be/hw-2inNNhHk>
- Gregory, S. (2021). *Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism.* Journalism. <https://doi.org/10.1177/14648849211060644>
- Gregory, S. (2021). *The world needs deepfake experts to stem this chaos.* WIRED. <https://www.wired.com/story/opinion-the-world-needs-deepfake-experts-to-stem-this-chaos/>
- Groh, M., Sankaranarayanan, A. y Picard, R. (2022). *Human detection of political deepfakes across transcripts, audio, and video.* ArXiv. <https://doi.org/10.48550/arXiv.2202.12883>
- Hart, A. (2021). *Voice AI is scary good now. Video game actors hate it.* INPUT Magazine. <https://www.inverse.com/input/gaming/video-game-voice-ai-human-actors-witcher-3-mod-controversy>
- Han, B. (2014). *Psicopolítica.* Herder.
- Hasan, H. R. & Salah, K. (2019). *Combating deepfake videos using blockchain and Smart Contracts.* IEEE Access. <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8668407>
- Hukkelås, H.; Mester, R.; Lindseth, F. (2019). *DeepPrivacy: A Generative Adversarial Network for Face Anonymization.* Norwegian University of Science and Technology. <https://doi.org/10.48550/arXiv.1909.04538>

- Hwang, T. (2020). *Deepfakes. A Grounded Threat Assessment*. Center for Security and Emerging Technology. <https://doi.org/10.51593/20190030>
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Mihaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,... Kuhn, J., Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education*. ArXiv. <https://doi.org/10.35542/osf.io/5er8f>
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN. NVIDIA. <https://doi.org/10.48550/arXiv.1912.04958>
- Knight, W. (2019). A new deepfake detection tool should keep world leaders safe – for now. MIT Technology Review. <https://www.technologyreview.com/2019/06/21/134815/a-new-deepfake-detection-tool-should-keep-world-leaders-safe-for-now>
- Lapowsky, I. (2019). *Here's what happens when news comes with a nutrition label*. WIRED. <https://www.wired.com/story/gallup-poll-fake-news-ratings/>
- Marr, B. (7 de marzo de 2022). Synthetic video and the future of deep fakes for business. Bernard Marr & Co. <https://bernardmarr.com/synthetic-video-and-the-future-of-deep-fakes-for-business/>
- Meneses, J. P. (2021). *Deepfakes and the 2020 US elections: what (did not) happen*. ArXiv. <https://doi.org/10.48550/arXiv.2101.09092>
- Millière, R. (2022). *Deep Learning and Synthetic Media*. ArXiv. <https://doi.org/10.48550/arXiv.2205.05764>

- MIT Technology Review [@techreview] (2019, septiembre 18). *This is the deepfake of @glichfield interviewing Vladimir Putin (wink nudge nudge)*. #EmTechMIT. <https://twitter.com/techreview/status/1174316798700924929?s=20>
- Mosley, T. (2019). *Perfect Deepfake Tech Could Arrive Sooner Than Expected*. WBUR. <https://www.wbur.org/hereandnow/2019/10/02/deepfake-technology>
- Nogués, G. (2018). *Pensar con otros: Una guía de supervivencia en tiempos de posverdad*. ABRE.
- Ovadya, A. (14 de junio de 2019). *Deepfakes Myths: Common Misconceptions About Synthetic Media*. Alliance for securing democracy. <https://securingdemocracy.gmfus.org/deepfake-myths-common-misconceptions-about-synthetic-media/>
- Ovadya, A. y Whittlestone, J. (2019). Reducing malicious use of synthetic media research: considerations and potential release practices for machine learning. ArXiv. <https://doi.org/10.48550/arXiv.1907.11274>
- Pinhanez, C. S., Flores, G. H., Vasconcelos, M. A., Qiao, M., Linck, N., de Paula, R. y Ong, Y. J. (2022). *Towards a New Science of Disinformation*. ArXiv. <https://doi.org/10.48550/arXiv.2204.01489>
- Polonski, V. (2017). *How artificial intelligence silently took over democracy*. World Economic Forum. <https://www.weforum.org/agenda/2017/08/artificial-intelligence-can-save-democracy-unless-it-destroys-it-first/>
- Puutio, A. (2020). *Deepfake democracy: Here's how modern elections could be decided by fake news*. World Economic Forum. <https://www.weforum.org/agenda/2020/10/deepfake-democracy-could-modern-elections-fall-prey-to-fiction/>
- Quach, K. (2023) *China follows through on plan to ban deepfake tech*. The Register. https://www.theregister.com/2023/01/10/china_deep_fake/?td=keepreading

- Radchenko, S. (2020). *Vladimir Putin Wants to Rewrite the History of World War II*. *Foreign Affairs*. <https://foreignpolicy.com/2020/01/21/vladimir-putin-wants-to-rewrite-the-history-of-world-war-ii/>
- Ratner, P. (2017). *Putin weighs in on Artificial Intelligence and Elon Musk Is Alarmed*. Big Think. <https://bigthink.com/politics-current-affairs/putin-weighs-in-on-artificial-intelligence-and-elon-musk-is-alarmed/>
- Rodríguez-Ferrándiz, R. (2019). *Posverdad y fake news en comunicación política: breve genealogía*. *El profesional de la información*, v. 28, n. 3. <https://doi.org/10.3145/epi.2019.may.14>
- Rowan, W. & Pears, N. (2022). *The effectiveness of temporal dependency in deepfake video detection*. University of York. <https://doi.org/10.48550/arXiv.2205.06684>
- Sharwood, S. (2022). *China bans deepfakes created without permission or for evil*. The Register. https://www.theregister.com/2022/12/12/china_deep_synthesis_deepfake_regulation/
- Tammekänd, J., Thomas, J. y Peterson, K. (2020). *Deepfakes 2020: The Tipping Point*. Sentinel. <https://thesentinel.ai/media/Deepfakes%202020:%20The%20Tipping%20Point,%20Sentinel.pdf>
- Thompson, N. (2019). Emmanuel Macron talks to WIRED about France's AI Strategy. WIRED. <https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy/>
- Toews, R. (2020). Deepfakes are going to wreak havoc on society. We are not prepared. Forbes. <https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=6e6c3ad87494>

- Valladão, A. G. A. (2018). *Artificial Intelligence and Political Science*. OCP.
<https://www.policycenter.ma/publications/artificial-intelligence-and-political-science>
- Wardle, C. y Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.
<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>
- Westerlund, M. (2019). *The Emergence of Deepfake Technology: A Review*. Technology Innovation Management Review. Vol. 9 Issue 11. (pp. 39-52)
<https://timreview.ca/article/1282>
- Zobaed, Z.; Rabby, M. F., Hossain, M. I., Hossain, E., Hasan, S., Karim, A. y Hasib, K. M. (2021). *Deepfakes: Detecting Forged and Synthetic Media Content Using Machine Learning*. ArXiv. <https://doi.org/10.48550/arXiv.2109.02874>