Facultad de Ciencias Humanas y Sociales
Grado en Relaciones Internacionales

Trabajo Fin de Grado

La respuesta de los poderes públicos en España ante la amenaza de los *Deepfake*

Estudiante: **Blanca Carcasona Prats**

Director: Emilio Sáenz-Francés

Madrid, junio 2019

“Hay dos maneras de dejarse engañar. Una es creer lo que no es cierto; la otra negarse a creer lo que es verdad”

Soren Kierkegard

Resumen

El fenómeno de las *fake news* ha adquirido gran relevancia en los últimos años y estamos asistiendo a nuevas variantes en las que la tecnología está haciendo que la distribución de contenidos falsos sea cada vez más difícil de detectar. Este es el caso de los *Deepfakes*, que permiten producir vídeos a través de programas de Inteligencia Artificial en los que se muestran imágenes y sonidos realistas de personas haciendo y diciendo cosas que nunca han pasado. En el siguiente trabajo se analizará el fenómeno de las *fake news* y de la era de la desinformación, junto con la irrupción de los *Deepfakes* y cómo se está afrontando este reto desde la comunidad internacional y los poderes públicos españoles. Finalmente, se analizarán una serie de cuestionarios a expertos en comunicación, política y ciberseguridad con el objetivo de averiguar cuáles son los potenciales riesgos a los que se enfrenta la sociedad española una vez esta tecnología se perfeccione, así como las posibles soluciones que se pueden llevar a cabo desde las esferas privadas, públicas y en colaboración con la Unión Europea.

Palabras clave

Deepfakes, *fake news*, desinformación, Inteligencia Artificial, riesgos, detección, prevención, ciberseguridad, estrategia, manipulación.

Abstract

The *fake news* phenomenon has become highly relevant lately and we are witnessing new variants in which technology is making the distribution of false content increasingly difficult to detect. This is the case of the *Deepfakes*, which produce videos through Artificial Intelligence programs that show realistic images and sounds of people doing and saying things that have never happened. The phenomenon of *fake news* and the era of disinformation will be analyzed, along with the emergence of *Deepfakes* and how this challenge is being addressed by the international community as well as the Spanish public powers. Finally, a series of questionnaires to experts in communication, politics and cybersecurity will be analyzed in order to find out what are the potential risks that the Spanish society faces once this technology is perfected, as well as the possible solutions that can be carried out by private and public spheres and in collaboration with the European Union.

Key words

Deepfakes, *fake news*, disinformation, Artificial Intelligence, risks, detection, prevention, cibersecurity, strategy, manipulation.

Índice

I. INTRODUCCIÓN

– *Finalidad y motivos*

Vivimos en una época en la que, gracias a la globalización, estamos interconectados en todos los niveles. En muchos casos esto es positivo, puesto que nos permite aprovechar las sinergias para conseguir objetivos comunes; sin embargo, al mismo tiempo esto hace que la población sea más vulnerable ante acciones hostiles. En este sentido, tanto la globalización, como el auge de las tecnologías y la creciente vulnerabilidad del mundo interconectado en el que vivimos, hace que surjan amenazas cada vez más difíciles de detectar, controlar y prevenir.

El objetivo del presente estudio es identificar tanto los riesgos como los mecanismos de defensa con los que cuenta el gobierno español para prevenir, detectar y reducir la amenaza de una nueva tecnología asociada a las *fake news* conocida como los *Deepfakes*. Dado que la temática es relativamente reciente, el grueso del estudio se centra en el análisis de la opinión de expertos en diferentes campos relacionados con los *Deepfakes* para averiguar las diferencias y similitudes que existen entre todos, respecto a la pregunta de investigación planteada:

¿Está preparado el gobierno español para hacer frente a la amenaza de los *Deepfakes*?

Los motivos por los que se ha elegido este tema, están relacionados con la evolución de la tecnología *Deepfake*, desde su uso con fines humorísticos, hasta su perfeccionamiento y empleo con fines políticos. La comunicación es el proceso de transmisión e intercambio de mensajes entre un emisor y un receptor, que se puede dar de manera escrita, visual y oral. A pesar de que la manipulación ha sido una práctica común en los medios de comunicación a lo largo de la historia, cada vez es más efectiva gracias a los avances en inteligencia artificial y recolección de datos. Los usuarios de internet generan su propio perfil socioeconómico y cultural a través del uso de diferentes plataformas, lo que permite a los medios de comunicación segmentar la información que desea ser enviada a diferentes grupos sociales, con el fin de generar enfrentamientos, especialmente en el campo de la política.

He hecho una revisión exhaustiva de la bibliografía principal del tema, que hasta día de hoy es algo limitada. Por tanto, el elemento que considero esencial en este trabajo es haber elaborado unos cuestionarios muy ambiciosos a personas expertas en campos relacionados con los *Deepfakes*. En circunstancias del COVID, es lo que proporciona un mayor valor añadido al trabajo. Además de esto, poder extraer conclusiones a partir de estos cuestionarios que permitan conocer no solo los riesgos presentes de esta tecnología, sino los riesgos futuros y sus posibles soluciones, pone en valor un posible marco de actuación que se podría llevar a cabo en España.

A nivel personal, de este trabajo me llevo varios aprendizajes personales, como por ejemplo la capacidad de adaptación a la hora de elaborar estos cuestionarios en el periodo actual de confinamiento. En segundo lugar, la capacidad de investigación y búsqueda de información sobre una temática muy reciente, que me ha permitido buscar diferentes recursos para tener una base bibliográfica sólida desde la que fundamentar el cuestionario.

Quiero dar las gracias a mis amigos y familiares cercanos, sin los que este trabajo no habría sido posible.

- *Objetivos*

El objetivo principal de este trabajo, como se ha comentado previamente, es averiguar si el gobierno español está preparado para hacer frente a la amenaza de los *Deepfakes* en la actualidad y en el futuro, cuando esta técnica de manipulación consiga perfeccionarse a niveles que sea imperceptible distinguir la realidad. Para ello, primero se ha hecho una revisión de la literatura en donde los objetivos que se persiguen son los siguientes:

- Exponer las principales características de la era de la desinformación
- Analizar el alcance y peligro del fenómeno *fake news*
- Presentar una definición de los *Deepfakes* y sus principales amenazas
- Evaluar la capacidad de respuesta actual ante esta amenaza por parte de la comunidad internacional en general, y de la Unión Europea en particular
- Explicar las diferencias en la gestión de la libertad de expresión y la limitación de lo publicado en Internet por parte de Estados Unidos y de la Unión Europea

Una vez elaborado un primer acercamiento respecto a esta tecnología, se ha elaborado un cuestionario de opinión a expertos en comunicación, política y ciberseguridad con los siguientes objetivos:

- Presentar los progresos hechos hasta ahora por parte del gobierno español en materia de ciberseguridad
- Averiguar si verdaderamente esta tecnología supone una amenaza real para la democracia, la credibilidad de los medios de comunicación y el intercambio efectivo de información entre países, entre otros
- Identificar qué papel tiene cada campo de estudio en la lucha contra esta amenaza
- Analizar los tipos de respuesta que debe dar el gobierno español para combatir esta amenaza

– *Metodología*

Para conseguir estos objetivos, se aplica la metodología deductiva, en donde primero se ha hecho una revisión de la literatura sobre la evolución y el estado actual de la tecnología *Deepfake*, para finalmente poder enmarcar adecuadamente las preguntas que más adelante se plantean en los tres cuestionarios de opinión a expertos. La conclusión se extrae de la mezcla entre lo que se ha hecho hasta ahora, junto con lo que los expertos esperan o desean que se haga en el futuro, con el objetivo de obtener respuestas a la forma en la que debe enfrentarse el gobierno español ante esta amenaza.

II. ESTADO DE LA CUESTIÓN

2.1. Desinformación: poder y manipulación en la era digital

La desinformación, según la Real Academia Española, es sinónimo de desconocimiento, confianza y manipulación. Este término se vincula cada vez más a el lenguaje periodístico y político, donde aparece como la manipulación de los medios de comunicación y el control de la información con fines políticos o económicos (Rodríguez Andrés, 2017).

Es de conocimiento común que a lo largo de la Historia diferentes actores han utilizado la manipulación y la mentira como herramientas para adquirir mayor poder e influencia. La sociedad ha asumido esto como normal, especialmente entre los políticos, lo que ha provocado un escenario en el que la verdad resulta imperceptible (Habermas, 2006).

La opinión pública tiene la certeza de que las nuevas tecnologías han hecho más vulnerable la verdad, al facilitar el acceso a nuevos formatos de manipulación y mentira (Torres et al., 2019). Sin embargo, existe controversia respecto a las formas en que se utilizan estas herramientas en las campañas de desinformación. Por esta razón, en la última década se han intensificado las campañas de distintos gobiernos para crear mayor resiliencia entre los ciudadanos frente a la desinformación (Casanova, 2018). En estas campañas, se empezó prestando especial atención a la forma en la que se construía el mensaje, descuidando por otro lado el público concreto al que debía ir dirigido (Blanco, 2018). No se tenía en cuenta la opinión preconcebida que podía tener el lector en base a sus experiencias, creencias y personalidad. Sin embargo, con el paso del tiempo se descubrió la razón de éxito de los agentes manipuladores: su habilidad para segmentar a su público objetivo en base a diferentes factores personales recopilados a través de internet (Garth y O'Donnell, 2006). Además, las personas construimos diferentes significados para un mismo mensaje, lo cual es clave para entender por qué la era digital ha hecho posible redirigir esos significados a aquello que el agente manipulador desea instalar entre nuestras creencias.

La propaganda tradicional, aunque conseguía persuadir a su público por ser el único medio de información existente, tenía el inconveniente de que se enfrentaba a un público

que asumía la información de manera pasiva, con lo cual la misma información se distribuía masivamente entre los distintos grupos de la población (Bertolin, 2017). Además, esto creaba un distanciamiento entre el lector y el creador de la información, generando incertidumbre y escepticismo ante su veracidad. A pesar de ello, este escepticismo ha aumentado en las últimas décadas tras la creación de internet, donde se han ido desarrollando nuevas herramientas tecnológicas de manipulación (El Diario, 2020). Estas tecnologías permiten segmentar a la población a través de la información que se publica, comparte, comenta y busca. De esta manera, todos los usuarios de internet contribuyen de manera inconsciente al proceso de información, por un lado, y de desinformación por otro. Según Torres, "somos actores en nuestro propio consumo de información" (2019, p. 7), lo que representa por primera vez en la Historia, un cambio muy importante en la lucha por obtener poder e influencia.

Según el Centro Criptológico Nacional (CCN) la metodología empleada en las campañas de desinformación digital se basa en las siguientes fases (2019, p.18):

a) **Análisis y detección de vulnerabilidades sociales y políticas de un país.** La estrategia para la alteración de la opinión pública no consiste en instaurar nuevas creencias entre los ciudadanos, sino más bien en generar confrontaciones ya existentes para conseguir una mayor polarización
b) **Creación de narrativas "transmedia".** Tras la detección de las temáticas que pueden generar debate y polarización en una sociedad, los agentes implicados buscan plataformas a través de las que distribuyen la información.
c) **Creación de una red de medios propios.** La estrategia para la distribución de contenidos digital se asienta bajo tres elementos: "medios propios, medios pagados y medios ganados". En este sentido, las narrativas disruptivas se introducen en el debate digital a través de una red de medios propios controlados por agentes de desinformación.
d) **Creación de canales de distribución automatizados.** El último paso consiste en la distribución directa de dicha información a través de cuentas automatizadas en redes sociales que se difunden masivamente y de forma segmentada a distintos grupos de la población.

2.2. El alcance y peligro del fenómeno *fake news*

Las noticias falsas o *fake news* son noticias emitidas por medios de comunicación, cuyo contenido es falso, pero a pesar de ello pueden llegar a ser aceptadas como creíbles y provocar una crisis de seguridad y confianza en el Estado (Centro Criptológico Nacional, 2019).

El fenómeno *fake news* o noticias falsas, no es un fenómeno nuevo, pero está adquiriendo cada vez mayor interés en los medios de comunicación por los efectos que tiene sobre el campo de la política. En los últimos meses, las expresiones *fake news* o *post-verdad* están adquiriendo una magnitud ampliamente reconocida (Maldonado, 2016). En 2016, el diccionario de Oxford reconoció la palabra *post-verdad* como palabra internacional del año, definiéndola como "aquellas circunstancias en que los hechos objetivos influyen menos en la formación de la opinión pública que las referencias a emociones y creencias personales" (Fernández-García, 2017, p.1). Esto quiere decir que cuando hablamos de *post-verdad*, nos estamos refiriendo a una falsedad que reconocemos como incierta pero que aceptamos como cierta, incluso utilizándola como base para tomar decisiones.

Dado que la manipulación en los medios de comunicación ha sido una práctica habitual empleada durante décadas por agentes públicos y privados, es importante averiguar cuál es el motivo que ha convertido a las *fake news* en un fenómeno mediático que supone un riesgo para el ámbito de las relaciones internacionales (Vallés, 2010). Por esta razón, se debe hacer un análisis sobre el sistema de acceso a la información en la era digital y sobre el papel de las plataformas digitales en su propagación.

Estas noticias falsas tienen aceptación entre el público y son fácilmente difundibles por las siguientes razones (Centro Criptológico Nacional, 2019):

- Contienen ciertos elementos verdaderos que las hacen verosímiles
- Están presentadas de manera sensacionalista para atraer la atención del lector
- Se difunden a través de medios de comunicación de reciente creación
- No hacen referencia a ninguna fuente de información
- A pesar de tener mayor repercusión en el corto plazo, pueden tener una gran influencia en el largo plazo. Esto es así, puesto que con el tiempo pueden ser

tomadas como referencia por medios de comunicación fiables y propagar la noticia de manera más creíble (Centro Criptológico Nacional, 2019).

2.3. ¿Qué son los *Deepfakes*?

Según Bazarkina, los *Deepfakes* son vídeos producidos a través de programas de Inteligencia Artificial (IA) en los que se muestran imágenes y sonidos realistas de personas haciendo y diciendo cosas que nunca han pasado. La tecnología *Deepfake* produce falsificaciones cada vez más difíciles de detectar, mediante el uso de herramientas de software que permiten imitar gestos, movimientos y expresiones faciales de la persona manipulada, así como de su voz a través de la entonación, el acento y el ritmo a el que habla el sujeto en cuestión (2019). Por otro lado, Villasenor define los *Deepfakes* como "una construcción creada con el objetivo de hacer que una persona diga o haga algo que nunca ha hecho o dicho" (2019, p.6).

La agencia principal de investigación del Departamento de Defensa de Estados Unidos, conocida como DARPA, está invirtiendo millones de dólares en crear herramientas de detección de *Deepfakes*; sin embargo, reconoce que si un agente utiliza Redes Generativas Antagónicas (GANs), en la mayoría de casos será imposible detectarlos. Las redes GANs son una serie de algoritmos de Inteligencia Artificial no supervisados, que funcionan a través del trabajo conjunto de dos redes, una que se encarga de crear el *Deepfake* y otra de evaluar su calidad (Schroeder, 2020).

El fenómeno *Deepfake* empezó a tener repercusión en 2007 cuando un usuario de Reddit, traspuso la cara de una celebridad sobre la de una actriz con fines pornográficos. Desde entonces, esta tecnología ha ido mejorando a través de la Inteligencia Artificial, que es capaz de comprimir cientos de vídeos e imágenes de una persona para más adelante, reproducir un vídeo hiperrealista en el que aparece esa persona diciendo o haciendo cosas que jamás han ocurrido (Vaccari y Chadwick, 2020). Esta tecnología "aprende" cómo actúa esa persona, es decir, recopila imágenes y vídeos en los que aparece la persona hablando y moviéndose y se apropia de sus gestos para crear vídeos que engañen al ojo del espectador. Esto se hace a través de técnicas de aprendizaje automático o *Machine Learning*, que permiten automatizar procesos que hasta entonces se habían hecho manualmente al editar vídeos a la velocidad de una máquina (Kumar, 2019).

El problema es que desde que aparecieron los *Deepfakes*, ahora es más simple, barato y rápido producirlos ya que se requieren menos datos, tiempo y potencia de cálculo para ofrecer imágenes de calidad. Sin embargo, a pesar de que es posible hacer *Deepfakes* rudimentarios a través del móvil con tecnologías sencillas, para crear *Deepfakes* verdaderamente creíbles, aún se requiere un software especializado aún por desarrollar.

2.4. Los *Deepfakes* y sus principales amenazas

Los *Deepfakes* no solo son una potencial herramienta para generar distorsiones de la realidad con fines humorísticos, sino que existen numerosos ejemplos en los que ya se ha utilizado esta tecnología en el campo de la política. Esta tecnología se ha empleado y puede llegar a emplearse en escenarios muy diferentes, por eso las amenazas a las que se enfrenta la comunidad internacional pueden resumirse en los siguientes puntos:

- **Desconfianza en la era de la desinformación**

El efecto que provoca la propagación de un *Deepfake* no solo depende de lo persuasivo que resulte el contenido de el vídeo, sino también de la credibilidad que tenga la fuente de información en la que se publica. Como se ha podido observar, la publicación de un vídeo inmediatamente anterior a unas elecciones puede generar una verdadera desconfianza entre los electores antes de que el vídeo pueda ser desacreditado.

Llegados a este punto, el sentido humano no es suficiente para identificar un fraude digital. La facilidad que existe para esconder acciones bajo un anónimo en Internet agrava el problema, puesto que, al existir libertad de expresión y, por ende, de divulgación de la información, es complicado distinguir entre lo que es verdadero y lo que es falso (Chesney, 2019). Tanto grupos extremistas como oponentes políticos autoritarios se encuentran entre los mayores beneficiarios de la "democratización del conocimiento y de la comunicación". En definitiva, los *Deepfakes* suponen una nueva herramienta para crear desconfianza y confusión entre la población de cualquier país y también a nivel global.

- **Amenaza directa a individuos**

La amenaza más inmediata de los *Deepfakes* va dirigida a individuos vulnerables. Esto puede ir dirigido a cualquier grupo de individuos, ya sean políticos, altos cargos directivos, mujeres o grupos étnicos particulares, entre otros.

Existen numerosos ejemplos que evidencian el alcance que tiene esta amenaza. Uno de ellos, es el estudio que se hizo en 2019 sobre 15.000 vídeos *Deepfake* encontrados en Internet, en donde se demostró que el 96% estaban hechos con fines pornográficos. En lo relativo al fraude, ese mismo año una red de informáticos estafó 243.000 dólares a una compañía energética británica tras publicar un vídeo de un alto directivo, manipulado a través de Inteligencia Artificial, en el que demandaba que se pagara de inmediato a un deudor de la compañía anónimo (Ohm, 2010). En cuanto a discriminación hacia grupos étnicos, está el caso del asesinato de ocho personas en la India en junio de 2018, tras la viralización de un vídeo a través de Whatsapp en el que aparecían como los autores de numerosos secuestros infantiles.

- **Creciente amenaza a las democracias**

Los *Deepfakes* pueden ser, y en efecto han sido usados para desprestigiar procesos democráticos. El *2019 Worldwide Threat Assessment* producido por la Comunidad de Inteligencia Estadounidense, advirtió de que "*adversarios y competidores estratégicos probablemente tratarán de utilizar Deepfakes o tecnologías de Machine Learning similares para crear imágenes, audios y vídeos convincentes pero falsos, con el objetivo de incrementar la influencia de campañas dirigidas en contra de los Estados Unidos y sus aliados"* (2019).

Los *Deepfakes* son una nueva alternativa ante un viejo y consolidado problema: la manipulación como herramienta política. Existen evidencias de que Joseph Stalin ordenara recortar a enemigos políticos de fotografías hace 80 años. Además, en 1990 la creación de Adobe Photoshop permitió que miles de usuarios tuvieran acceso a la edición y manipulación de imágenes. En ese mismo año, el periódico americano *The Newsweek* advirtió del peligro que esto podía causar, ya que gobiernos autoritarios como el de China podían utilizar estas herramientas para ocultar futuras atrocidades como la de las protestas

en la plaza de Tiananmen en 1989, ya que la fotografía electrónica, al convertirse en maleable, hace posible negar su veracidad (Stockler, 2019).

Esta información engañosa toma especial fuerza en comunidades altamente fragmentadas, donde los ciudadanos rara vez se enfrentan a ideas que desafíen sus propias creencias (Schroeder, 2020). Es en este tipo de sociedades donde la información engañosa que apoya narrativas dominantes es más propicia a aceptarse como verdadera y difundirse con mayor rapidez.

2.5.¿Cómo está reaccionando la comunidad internacional?

En agosto de 2019 unos investigadores israelíes anunciaron una nueva técnica para elaborar *Deepfakes* que consiste en superponer una cara a la de otra persona que está hablando. A diferencia de otros métodos, con esta técnica se pueden manipular vídeos donde hay más cosas sucediendo y la persona no es el centro de atención de la escena. Esto dota a los vídeos de un incluso mayor hiperrealismo, puesto que la manipulación es prácticamente imperceptible y el tiempo de elaboración es mucho más reducido (Galston, 2020). Además, dado que han anunciado el proceso de elaboración públicamente, la proliferación de esta barata y sencilla tecnología parece inevitable.

Antes de que se desarrollara esta avanzada tecnología *Deepfake*, se denunciaron numerosos casos en los que la divulgación de vídeos manipulados había resultado creíble y profundamente dañina (Harwell, 2019). Por ejemplo, en 2017, un grupo británico de extrema derecha publicó en Twitter un vídeo de unas personas musulmanas celebrando en Londres el ataque terrorista de Bataclán. Más tarde se descubrió que se habían utilizado imágenes de un grupo de pakistaníes que en 2009, habían estado celebrando la victoria de Pakistán en la Copa Mundial de Cricket. Por otro lado, en 2018, una empresa de medios de comunicación estadounidense llamada *Buzzfeed*, publicó un vídeo manipulado en el que aparecía Barack Obama insultando a Donald Trump. La intención del vídeo era denunciar el peligro de creer toda información que aparece en Internet, así como de alertar de lo sencillo que es crear este tipo de vídeos. En mayo de 2019 se publicó un vídeo en el que aparecía la Presidenta de la Cámara de Representantes de Estados Unidos, Nancy Pelosi, en un evidente estado de embriaguez. Donald Trump compartió este vídeo a través de Twitter, donde millones de usuarios lo comentaron y compartieron

(Venkatararamakrishnan, 2019). Más tarde se descubrió que la voz de Pelosy había sido manipulada a través de la tecnología *Deepfake*, pero el daño a su reputación ya estaba hecho, puesto que mucha gente tampoco se creía que el vídeo hubiera estado manipulado.

En este sentido, el fenómeno *Deepfake* no sólo es utilizado para difundir una mentira, sino también para aprovechar ese ambiente de engaño e incertidumbre para defender como falsa una información verdadera (Colomina, 2019). Ejemplo de ello, es el uso que hizo de este un teórico conspiracionista llamado Alex Jones, quien denunció a través de diferentes redes sociales que las imágenes sobre el tiroteo de Sandy Hook en Estados Unidos, en el que murieron 20 jóvenes, era un montaje creado por los detractores del uso de armas en Estados Unidos.

Como se puede observar, no solo el uso de *Deepfakes* con intención maliciosa viene del lado de transformar la mentira en verdad, sino también de hacer que la verdad parezca mentira. Según *The New York Times*, a estos efectos se les conoce como *The Oxygen Theory* y *The Liar's Dividend*. El primero, se refiere al efecto que se genera cuando una noticia es falsa y se la desacredita, ya que se le da una mayor repercusión y "vida" de cara al público (Metz, 2019). Finalmente, la veracidad de la noticia deja de tener importancia porque la fuente queda desacreditada. *The Liar's Dividend* en cambio, se refiere al efecto que se produce cuando alguien utiliza un vídeo que es real y dice que es falso.

Este problema parece especialmente alarmante a medida que se aproxima el periodo de elecciones presidenciales en Estados Unidos. Si se está llegando a un punto en que la Inteligencia Artificial hace virtualmente imposible detectar representaciones de sonido y vídeo, ver no va a volver a significar creer y cada individuo tendrá que decidir por su cuenta qué es lo que decide creer y lo que no (Knight, 2019). Por tanto, resulta importante combatir el acelerado ritmo al que se desarrollan los mecanismos que hacen cada vez más creíbles los *Deepfakes*.

2.6. ¿Es capaz la Unión Europea de evitar que los *Deepfakes* supongan una amenaza para la paz?

Según EU-Listco, un portal que investiga los desafíos a los que se enfrenta la política exterior europea, a pesar de que el uso de *Deepfakes* ya es accesible a el público a través de aplicaciones de móvil como FaceApp o Zao, la Unión Europea aún no ha abordado el

problema con contundencia. En su plan de acción contra la desinformación, reconoce que "la desinformación es una herramienta de influencia poderosa y muy barata" (Bressan, 2019, p.1), sin embargo, no especifica qué acciones se deben llevar a cabo para prevenir a los europeos del daño potencial que les puede causar.

Es cierto que la Unión Europea lanzó en 2015 el proyecto "EU vs Disinfo" con el objetivo de estimar, identificar y responder a las campañas de desinformación publicadas por la Federación Rusa en contra de la Unión Europea y sus estados miembros. Con este proyecto se pretendía dar visibilidad a las operaciones del Kremlin para hacer a los ciudadanos europeos más críticos y resilientes frente a los ataques de contenido digital por parte de Rusia (European Union, 2015). Sin embargo, el proyecto recibió un gran volumen de críticas debido a la falta de recursos que se pusieron a su disposición. Este ha sido el único proyecto que se ha llevado a cabo por parte de la Unión Europea para combatir la desinformación hasta la fecha.

El 25 de marzo de 2019, el portal EU-Listco hizo un ejercicio de prevención sobre tecnologías no gobernadas, en donde una serie de expertos identificaron dentro de esta categoría tecnologías sobrevaloradas e infravaloradas y elaboraron diferentes escenarios en los que estas pudieran aumentar el riesgo de conflicto y quiebra del sistema democrático en Europa (Bressan, 2019). Entre los escenarios desarrollados, los *Deepfakes* destacaron como una de las tecnologías con mayor potencial de riesgo. Los factores de riesgo que pueden desencadenar en violencia que se identificaron fueron, entre otros, las sociedades polarizadas, la marginalización de grupos y una historia común de conflicto. La combinación de estos factores en situaciones volátiles, como puede ser la celebración de elecciones en un país, tiene un particular potencial de riesgo.

De este taller, el portal concluyó en que es necesaria una respuesta política europea más robusta y una mayor iniciativa de los estados miembros para aunar esfuerzos en la prevención estructural, efectiva y a largo plazo de los daños que la desinformación puede ocasionar. Para ello, la inversión debe ir destinada a la educación y concienciación de los ciudadanos, además de al desarrollo de habilidades tecnológicas que combatan la distribución de los *Deepfakes* (EU-Listco, 2019). Todo ello fomentando la libertad de expresión de información veraz en los medios de comunicación, con el objetivo de recuperar la confianza de los ciudadanos en sus gobiernos y entre diferentes comunidades.

Incluso se plantean soluciones como la de que los políticos y otros cargos públicos lleven un equipo de registro móvil continuo, junto con "servicios de autentificación, parecidos a las cámaras que en ocasiones portan consigo los policías", para ser capaces de demostrar siempre que sus alegaciones son reales en caso de ser víctimas de un vídeo manipulado a través de la tecnología *Deepfake* (Bressan, 2019). Sin embargo, esta parece una medida extrema que en muchas ocasiones podría atentar contra la propia intimidad de los individuos.

A pesar de las dificultades que supone recibir apoyo financiero y político a largo plazo, la Unión Europea debe ser consciente de que tecnologías como los *Deepfakes* están distribuyendo información falsa desde esferas de rivalidad geopolítica, con lo cual la respuesta que se debe dar, para lograr ser efectiva, tiene que ser a nivel comunitario.

III. MARCO NORMATIVO

Aunque la distribución de información engañosa no es algo nuevo, es cierto que, con estas nuevas amenazas a la verdad, los gobiernos encuentran cada vez más dificultades para limitar el uso y distribución de *Deepfakes* y respetar el derecho de libre expresión de los ciudadanos. De hecho, en Estados Unidos y la Unión Europea, que son las dos regiones donde se está haciendo mayor esfuerzo por frenar esta amenaza, se han tomado enfoques muy diferentes al respecto (Schroeder, 2020). La limitación del uso de *Deepfakes* sobre el discurso democrático tiene distinta intensidad en uno y en otro, y radica de las distintas interpretaciones que se dan sobre el nivel de intervencionismo que debería tener el gobierno sobre la difusión de información.

3.1. Marco legal de Estados Unidos para la salvaguardia de información engañosa

La base de la ideología americana a este respecto, está asociada a interpretaciones libertarias de ideas de la Ilustración sobre la naturaleza de la verdad y la racionalidad del individuo. John Milton explicaba que allá donde hay deseo por aprender, habrá necesidad de discutir y escribir sobre las diferentes opiniones que de ello deriven. Años más tarde, tras un caso de difamación, Luis Powell alegó ante la Corte Suprema de Justicia Estadounidense que *no existe tal cosa como una idea falsa* (Schroeder, 2020, p.5). Así que, para evitar el problema persistente que se derivaba de la limitación al derecho de libre expresión, establecido en la Primera Enmienda a la Constitución de Estados Unidos, el marco legal americano comenzó a adoptar un enfoque de menor intervencionismo gubernamental (Kahn, 2020).

3.2. Marco legal de La Unión Europea para la salvaguardia de información engañosa

Complementariamente, en el marco legal europeo se han adoptado papeles más paternalistas. En 1998, en el caso de Hertel v. Switzerland en la Corte Europea de Derechos Humanos se reconoció que “la libertad de expresión constituye una de las fundaciones esenciales de una sociedad democrática y una de las condiciones básicas para su progreso y para la autorrealización de cada individuo” (Schroeder, 2020, p.5). A pesar de ello, la Corte también reconoció que hay una serie de limitaciones a la libertad de

expresión, considerando si la información que el gobierno restringe es necesaria en una sociedad democrática o no.

De acuerdo con esta línea de pensamiento, en un informe publicado en 2019 por el Parlamento Europeo se reconoce que "la tecnología *Deepfake* presenta un problema incluso más difícil de gestionar que la manipulación de contenido textual, puesto que es una herramienta más propensa a desencadenar emociones más fuertes" (European Parliament, 2019). Esta conexión más visceral con la información, junto con los estudios que demuestran que la información transmitida por vídeo es más fácil de recordar, presentan un escenario de preocupación respecto al tipo de información a la que los individuos cada vez tendrán mayor acceso (Sánchez, 2019).

IV. ANÁLISIS DE LA ESTRATEGIA DE ESPAÑA ANTE LA AMENAZA DE LOS *DEEPFAKES*

4.1. Estrategia Nacional de Ciberseguridad 2019

La Estrategia Nacional de Ciberseguridad se desarrolla con la intención de dictaminar los objetivos propuestos por el gobierno español, así como las líneas de acción marcadas para conseguirlos.

Principales riesgos:

Según este documento publicado por el gobierno de España en 2019, hay más de 20 millones de españoles en riesgo de ser víctimas de este tipo de acciones maliciosas (Consejo de Seguridad Nacional, 2019). Esto es así, porque existen 27,6 millones de usuarios en España, de los cuales 25,5 hacen uso diario de internet y redes sociales. De acuerdo con los datos obtenidos por el Observatorio Nacional de las Telecomunicaciones, el 92% de la población española entre 16 y 65 años utiliza internet sistemáticamente para obtener información y el 85% toma como fuente de información principal las redes sociales. Además, un estudio elaborado en 2017 por la Asociación para la Investigación de Medios de Comunicación (AIMC) manifiesta que el 84,6% de los usuarios españoles utilizan internet para leer noticias de actualidad. Con lo cual, los resultados obtenidos por ambos estudios revelan que aproximadamente el 90% de los ciudadanos españoles, de edades entre 16 y 65 años son víctimas potenciales de acciones de desinformación (Centro Criptológico Nacional, 2019). Un ejemplo del nivel de ignorancia que hay respecto a la generación de noticias por plataformas digitales, es que "apenas 3 de cada 10 usuarios de internet son conscientes de que las noticias que leen en Facebook dependen de un algoritmo" (Universidad de Navarra, 2018, p.14). Este nivel de desconocimiento sobre la información que recibimos de internet, constituye un riesgo más para la vulnerabilidad de la opinión pública en España.

Es por eso que el Decálogo de Buenas Prácticas sobre Desinformación en el Ciberespacio elaborado por el Centro Criptológico Nacional incluye de manera explícita la amenaza de estos actores como uno de los retos de seguridad prioritarios a los que debe hacer frente el país. El origen de estas amenazas radica en la compleja realidad del panorama español, tanto en el ámbito sociopolítico como en el económico (Centro Criptológico Nacional,

2019). Además, el mayor problema de estas acciones puede provenir tanto de agentes estatales como no estatales, con lo cual el foco del problema exige una especial atención. Estas acciones se denominan híbridas, porque "combinan el uso de medios militares con ataques cibernéticos, elementos de presión económica o campañas de influencia en las redes sociales." (Gobierno de Aragón, 2019, p. 1).

- **Pérdida de confianza en las redes sociales y los medios de comunicación**

Antes de la expansión masiva de internet, los medios de comunicación tradicionales eran los encargados de conformar la opinión pública y como única forma de comunicación, se identificaban como agentes de información absolutamente veraz. Los medios de comunicación instigaban los debates públicos y, en consecuencia, eran corresponsables de la cohesión social del Estado (Centro Criptológico Nacional, 2019).

Sin embargo, hoy en día la realidad es diferente. Según un estudio realizado por el *Edelman Trust Barometer* (2018), tan solo el 44% de los ciudadanos españoles que consumen internet confían en los medios de comunicación. Pero esto no es un caso aislado, puesto que, en una comparativa con otros países, los europeos en general muestran un alto nivel de desconfianza, siendo Turquía, Australia y Japón los países con más bajos porcentajes. Tan solo China, India e Indonesia muestran niveles de confianza de los medios de comunicación superiores al 60% de la población.

Debido a este alto nivel de desconfianza, un estudio realizado por la Universidad de Navarra y la Universidad de Oxford (2018) manifiesta que España es uno de los países con mayor nivel de preocupación por ser víctimas de campañas de desinformación. El 69% de los usuarios de internet en España reconoce sentirse preocupado por la abrumadora cantidad de información en internet, en muchos casos contradictoria, que les hace no saber diferenciar entre lo que es verdadero y lo que es falso. Como se puede observar, Brasil es el país con mayor nivel de preocupación con un 85%, seguido por Portugal con un 71% e inmediatamente después por España (Centro Criptológico Nacional, 2019).

- **Pérdida de confianza en la capacidad del ciudadano para actuar deliberadamente**

Algunas estrategias de desinformación buscan desestabilizar la confianza del ciudadano en lo más profundo del sistema político que conforma el país: *la confianza en las democracias de que el ciudadano mejor informado será capaz de tomar las mejores decisiones para su gobernanza* (Centro Criptológico Nacional, 2019, p. 15). En definitiva, buscan quebrar la confianza que tiene el ciudadano sobre su propia capacidad de decisión, es decir, sobre su habilidad para actuar democráticamente. Estas estrategias hacen que el ciudadano dude de sí mismo y, en consecuencia, piense que no puede actuar con autonomía ya que toda la información que recibe puede que esté manipulada para hacerle pensar de una manera concreta.

De esta manera, si se consigue generar un entorno de desconfianza hacia los medios de comunicación y las instituciones públicas, el usuario acabará teniendo la sensación de que no es capaz de establecer un juicio de valor adecuado para tomar decisiones de voto. Esta estrategia es comúnmente utilizada por países extranjeros, con el objetivo de distorsionar el sentido del voto en las democracias. Por ejemplo, si un Estado extranjero quiere influir sobre la credibilidad del sistema de voto en otro país, si logra acaparar titulares que así lo reconozcan, el ciudadano del país de origen acabará creyendo que tiene menos posibilidades de votar legítimamente, puesto que hay un país extranjero que le está manipulando (European Commission, 2018). Con lo cual, el reconocimiento de la intención maliciosa del Estado extranjero en los medios, le otorga legitimidad para ejercer esa influencia.

En España en concreto esto supone una especial amenaza, porque como se ha mostrado anteriormente, la población española es una de las poblaciones que más desconfianza y preocupación muestran ante la veracidad de la información que aparece en sus medios de comunicación. Según el Decálogo de Buenas Prácticas sobre Desinformación en el Ciberespacio elaborado por el Centro Criptológico Nacional, para combatir estas campañas de desinformación y conseguir anular el efecto generado, se deben llevar a cabo las siguientes acciones:

a) Identificar lo antes posible la intención maliciosa de posibles agentes interesados en la desestabilización del sistema político del Estado
b) Evaluarlos para poder hacer una distinción rápida y clara de los agentes legítimos y los ilegítimos

c) Contextualizar las intenciones y la capacidad de los agentes para analizar la dimensión de la manipulación en el contexto global actual
d) Crear campañas de información para el usuario con el objetivo de que este tenga un mayor conocimiento de los riesgos de manipulación que existen en internet, logrando mayor credibilidad entre los ciudadanos
e) Introducir el Decálogo de Buenas Prácticas tanto en estudios académicos como cualquier otro tipo de medio informativo (Centro Criptológico Nacional, 2019)

- **Pérdida de confianza en las instituciones públicas**

Ante este escenario de sobreexposición de la información y dificultad para discernir la veracidad de las noticias que se encuentran en internet, los responsables de las campañas de desinformación se aprovechan de este entorno de incertidumbre generado, para también generar desconfianza en las instituciones públicas (European Commission, 2018). El informe realizado por el *Edelman Trust Barometer* manifiesta que, en España, tan solo el 24% de los cibernautas confía en sus instituciones públicas.

Este es precisamente el objetivo de estos agentes: detectar las vulnerabilidades del contexto socioeconómico y político de cada país, y potenciarlos para aumentar el clima de desconfianza (Centro Criptológico Nacional, 2019).

- **Polarización social**

Las plataformas digitales de información utilizan algoritmos para seleccionar noticias de manera personalizada al usuario. De esta manera, la propia naturaleza de estos medios genera una polarización de la sociedad, reforzando las creencias de los individuos y acotando los espacios de debate. Un ejemplo ilustrador de esta situación, son los descubrimientos que hizo un estudio elaborado en 2017 por el *MIT Social Machines Laboratory* durante la campaña electoral estadounidense de 2016. En este estudio, se detectaron altos niveles de polarización entre las comunidades de votantes de Donald Trump y Hillary Clinton. Debido a esto, el debate entre ambos grupos fue prácticamente inexistente, puesto que el poco que hubo, solo logró radicalizar las posturas de ambas comunidades (Meyer, 2018).

Del mismo modo, un estudio elaborado por *Alto Analytics* descubrió que el debate digital sobre la inmigración en Italia con el ascenso al poder de Matteo Salvini, también estaba altamente polarizado y ambas posturas apenas se relacionaban entre sí (Alto Analytics, 2017). Además, se presentaron evidencias en el estudio de que pudo haber interferencias de medios extranjeros y otros agentes, que intervinieron con el objetivo de aumentar las diferencias de opinión.

En España, esto también ha ocurrido numerosas veces en noticias que enfrentan a partidos de derechas con partidos de izquierdas, así como en noticias respecto al independentismo catalán. En octubre de 2017, la revista rusa *RT News* publicó una noticia en la que afirmaba que el gobierno español había mandado tanques militares a Barcelona para frenar las protestas en Cataluña sobre el independentismo. A pesar de la credibilidad de la fuente que lo había publicado y de que dicha noticia no estaba contrastada con ninguna imagen, la noticia fue compartida a través de Facebook por 11.800 usuarios. En este caso no se emplearon imágenes ni vídeos manipulados, ya que bastó con un titular atractivo y la supuesta "validación" de la información de la mano de analistas políticos (RT News, 2017). Sin embargo, tras ver el alcance de una noticia así, resulta importante cuestionarse cuál habría sido entonces el alcance que habría tenido una noticia con este titular si la revista se hubiera apoyado de vídeos manipulados con la tecnología *Deepfake*.

Líneas de acción y medidas:

Según el informe de la Presidencia del Gobierno respecto a la Estrategia Nacional de Ciberseguridad de 2019, España es uno de los países más interconectados del mundo. Esto, unido a la peligrosidad de las amenazas cibernéticas a las que se enfrentan los ciudadanos, hace necesaria la alineación de este plan estratégico con la Estrategia de Seguridad Nacional de 2017. Según lo estipulado, el objetivo es *hacer de este momento de cambios, no una fuente de malestar cultural y de regresión económica y laboral, sino una oportunidad para incrementar la competitividad de España y el bienestar de los españoles y las españolas junto a la de nuestros socios europeos* (Administración General del Estado, 2019, p.3). Debido a la transversalidad de este tipo de amenaza, el informe reconoce la necesidad de cooperar de manera global y actuar de acuerdo con el Derecho Internacional, en base a los principios recogidos en la Constitución Española y en la Carta de las Naciones Unidas.

La primera Estrategia Nacional de Ciberseguridad se publicó en 2013 para fijar las directrices a seguir y diseñar el modelo de gobernanza de seguridad nacional respecto al espacio cibernético. Tras este plan se creó en 2014 el Consejo Nacional de Ciberseguridad como rama de apoyo técnico al Consejo Nacional de Seguridad, con el objetivo de coordinar los organismos especializados en ciberseguridad nacional.

Con respecto al marco legal sobre ciberseguridad, en 2015 se modificó el Esquema Nacional de Seguridad con el objetivo de proteger al Sector Público de ataques cibernéticos. Ese mismo año entró en vigor la Ley 36/2015, relativa a la Seguridad Nacional, en donde se reconoció la ciberseguridad como uno de los ámbitos de mayor interés para la defensa del país. Más tarde, se publicó la Estrategia de Seguridad Nacional 2017 con intención de hacer visible la necesidad de que la ciberseguridad ocupe un espacio propio y prioritario en el campo de la seguridad nacional (Centro Criptológico Nacional, 2019). De esta manera, en 2018 entró en vigor el Real Decreto-Ley de seguridad de las redes y sistemas de información, para garantizar la protección cibernética en todos los sectores estratégicos del país.

Plan estratégico:

Entre las amenazas que se presentan en el ciberespacio, destaca el ciber-espionaje, la ciber-criminalidad, el hacktivismo y las amenazas híbridas. Los *Deepfakes*, entendidos como una forma de aplicar las *fake news,* se encuentran dentro de la modalidad de amenazas híbridas, ya que estas amenazas se entienden como la combinación de ataques militares con campañas de desinformación dirigidas a atacar las vulnerabilidades de los sistemas democráticos y sus instituciones. Por otra parte, el uso de datos personales que circulan por el ciberespacio facilita la efectividad de todas estas acciones, puesto que permite conocer los públicos objetivos donde se va a lograr un mayor impacto.

Los principios rectores a los que se somete la Estrategia de Ciberseguridad Nacional 2019 son la resiliencia, la eficiencia, la anticipación y la unidad de acción. En el diseño del plan se establecen cinco objetivos con sus consecuentes líneas de acción y medidas a implementar para hacer efectiva la estrategia:

a) ***Consolidar un marco legal que garantice la seguridad y resiliencia en los sistemas de información***. Para ello será necesario reforzar las capacidades de

detección, análisis, ciber-defensa y ciber-inteligencia del ciberespacio. Además, se deberán desarrollar instrumentos de prevención de amenazas a través de la colaboración de los centros de investigación con los sectores público y privado. También será necesario reforzar el marco legislativo para hacerlo más extenso y coherente, para asegurar la protección de las infraestructuras tecnológicas.

b) ***Fortalecer la cooperación policial y judicial*** para asegurar el uso seguro y fiable del ciberespacio. Para ello es necesario reforzar las capacidades de identificación y persecución de la ciber-criminalidad. En este sentido, la participación ciudadana en la transmisión de información de interés es fundamental. También es necesario agilizar el envío de la documentación necesaria a la jurisdicción penal y proporcionar a los profesionales de ambos ámbitos con los recursos técnicos y humanos necesarios.

c) ***La defensa del sector privado y de los ciudadanos,*** a través del intercambio constante de información sobre los riesgos del ciberespacio y el desarrollo de mecanismos de seguimiento e identificación de ciber-amenazas. Para ello es necesario ofrecer un servicio público de ciberseguridad fácil de acceder, además de impulsar el uso de ciberseguridad en las empresas y crear métodos de denuncia efectivos y rápidos.

d) ***Lograr autonomía tecnológica nacional*** para potenciar la cultura de la ciberseguridad y el compromiso ciudadano. Para ello es necesario contar con recursos técnicos y humanos mediante el desarrollo de actividades de I+D+i en ciberseguridad. También se propone crear un foro Nacional de Ciberseguridad en el que diferentes expertos puedan responder a preguntas de ciudadanos y entidades. En definitiva, lo que se busca es desarrollar una cultura de ciberseguridad para crear una concienciación ciudadana.

e) ***Contribuir a la creación de un marco internacional que garantice la ciberseguridad***. España considera que es necesario colaborar con las Naciones Unidas en la construcción de consensos y con la Unión Europea para lograr autonomía estratégica europea (Administración General del Estado, 2019).

Además, se busca reforzar la cooperación bilateral y poner especial atención en jóvenes y mujeres, que son los grupos de civiles de mayor riesgo en España.

4.2. Cuestionario de opinión a expertos

Dado que la creación de *Deepfakes* es relativamente reciente, se ha realizado un cuestionario de opinión a expertos en comunicación, política y ciberseguridad, con el objetivo de identificar los posibles riesgos y tipos de respuesta que se deben dar ante la amenaza de esta tecnología. El cuestionario contiene respuestas específicas a cada campo profesional, sin embargo, en todos ellos está incluida la pregunta principal de este estudio: ¿cómo se debe enfrentar España ante la amenaza de los *Deepfakes*?

4.2.1. Cuestionario a expertos en comunicación

Para este apartado, se ha entrevistado a Alberto Artero (Director General del periódico El Confidencial), Enrique Campo (Subdirector de Informativos de la COPE) y a Borja Adsuara (experto en Derecho, Estrategia y Comunicación Digital).

a) ¿Cuál es el riesgo que asume la sociedad ante el fenómeno de los Deepfake?

Según comenta Alberto Artero, es más importante la causa que la consecuencia. Sólo analizando la primera nos permitirá contestar la segunda, esto es: los riesgos. Lo que buscan los *Deepfakes* es la manipulación de las opiniones y/o percepciones de una realidad dotando de apariencia de realidad a lo que no son sino bulos. Para él, el mayor riesgo es la toma de decisiones inadecuadas por vivir en una sociedad desinformada y cada vez más fragmentada, en donde el aborregamiento social nos lleva a aceptar cualquier realidad. El ejemplo más claro de esto es lo que hacía Facebook con Cambridge Analytics para las campañas electorales: sin meter bulos alimentaban el rechazo a una opción política, alimentando otra. En línea con esto, Borja Adsuara cree que los Deepakes suponen una alternativa más para la falsificación de pruebas en juicios y para las campañas de desprestigio, entre otros.

Sin embargo, Enrique Campo opina que no existe riesgo alguno. Para asumir un riesgo tienes que conocer la amenaza, y la sociedad no conoce ni valora el fenómeno de los Deepfake. Sí lo conocen los expertos en tecnología y comunicaciones, pero tampoco ellos son capaces de saber hasta dónde puede llegar el uso de estas técnicas y sus

consecuencias. Afirma que 8 de cada 10 periodistas de cualquier medio de comunicación generalista español, desconocen lo que son los *Deepfakes*. La sociedad no asume ningún riesgo porque no conoce la amenaza. Por tanto, el riesgo real que se afronta es el de que el fenómeno se perfeccione, se generalice y llegue un momento en el que sea complicado distinguir un vídeo o audio manipulado de uno auténtico. Si eso ocurre, los medios de comunicación tendrán que ser aún más escrupulosos en la verificación de noticias. Sin embargo, lo que ocurra y corra como la pólvora por redes sociales como WhatsApp o Telegram será difícilmente desactivable.

b) ¿Por qué compensan las noticias falsas a quien se dedica a difundirlas?

Para Alberto Artero y Enrique Campo, cualquier iniciativa de este tipo tiene tres objetivos principales:

- **Social**, de movilización de masas y creación de estados de opinión favorables o contrarios a algo. La intención es desinformar, difundiendo falsedades y haciendo ruido mediático en las redes sociales para crear incertidumbre, inseguridad general o un estado de ánimo y de opinión determinado. Este es el caso de quienes difunden noticias falsas para desestabilizar relaciones entre potencias o entre gobernantes. Aún queda mucho por saber acerca de las injerencias rusas en procesos electorales europeos o norteamericanos.
- **De poder**, de ocupación de espacios de agentes nuevos. La intención es desprestigiar a una persona, un colectivo o una empresa. Encender la mecha es muy fácil en Internet; apagarla, complicado.
- **Económico**, por hacer negocio obteniendo visitas a una web o perfil social, logrando notoriedad gracias al tráfico redirigido a partir de la difusión de una noticia falsa. Otra forma de hacer negocio a través de las *fake news* en general, es directamente la suplantación de identidades para engañar y obtener datos o dinero. Es el caso de quienes imitan a bancos o a organismos oficiales, haciéndose pasar por ellos y pidiendo por correo datos personales o, directamente, transferencias económicas.
- **Diversión,** la intención es difundir noticias falsas por enredar, por entretenimiento, por ver qué pasa, plantear un desafío, ver hasta dónde llegó...

Por otro lado, Borja Adsuara analiza los objetivos en función del agente que los realiza, puesto que dice que hay potencias que buscan desestabilizar países, inversores que

quieren hacer tambalear una empresa cotizada y ciudadanos que disfrutan engañando a gente, entre otros.

c) ¿Cuál es el papel que deben tener las plataformas digitales en las que se divulga información respecto a la lucha contra los *Deepfakes*?

Los tres entrevistados coinciden en que las plataformas digitales deberían actuar activamente en la lucha contra los *Deepfakes*. Borja Adsuara cree que se deberían utilizar redes de bots para que, a base de repetición, pueda parecer verosímil enfrentarse a la viralización artificial de contenido falso. Por otro lado, Enrique Campo considera que no es lo mismo publicar contenido en el que se explica el proceso por el que se crea un *Deepfake*, y otra que se permita publicar tutoriales enseñando cómo hacerlos directamente.

Tanto Enrique Campo como Alberto Artero, afirman que al igual que hay una tecnología para crearlos, pronto habrá una tecnología para detectarlos, filtrarlos e impedir su publicación. Sin embargo, en la actualidad, los recursos necesarios para hacerlo con precisión y rapidez son tan escasos que en muchas ocasiones la detección no compensa financieramente. Además, Alberto Artero añade que, en las redes sociales, que no son medios de comunicación per se, cualquiera es libre de colgar el contenido que desee y en la mayoría de casos, no hay responsabilidad. Por esta razón, esa frontera que existe entre el creador de contenido y el distribuidor del mismo, es lo que hace que la responsabilidad que se puede exigir a las plataformas digitales sea limitada.

d) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de *Deepfakes* sin condicionar la libertad de expresión?

Igualmente, los tres coinciden en que el equilibrio entre ambos es complicado y controvertido. A pesar de ello, Enrique Campo subraya que el derecho a la libertad de expresión no ampara la comisión de delitos o engaños. El derecho a la libertad de expresión no es un derecho absoluto, sino que está regulado y tiene límites. En línea con esto, es perfectamente posible establecer un marco legal que prohíba el uso malicioso de *Deepfakes*. Algunos Estados norteamericanos ya lo han hecho. China también lo ha hecho, aunque, efectivamente, su marco jurídico no es homologable al de las democracias modernas. El uso y difusión maliciosa de *Deepfakes* podría ser incluido sin problemas en

el Código Penal español, como ya lo están otras prácticas que se desarrollan en el terreno virtual de Internet, quizás no tanto por la práctica concreta sino por el delito que se comete a través de ella, previamente tipificado y regulado. Es decir, aun no incluyendo expresamente las *Deepfakes* en el Código Penal, ya podrían ser perseguibles legalmente en los casos en los que se observara un delito por injurias, calumnia, contra el honor o contra la propia imagen, por ejemplo. Conjuntamente, Borja Adsuara añade que el único límite a la libertad de expresión es la ley.

Para Alberto Artero, la única manera es dotando de cualidades periodísticas a las informaciones, de tal manera que haya una identificación de fuentes solventes, un autor reconocido que responda de la veracidad de forma subyacente a un *nickname* y una responsabilidad económica en caso de irresponsabilidad. Lo idóneo sería que la distribución de contenidos por parte de las plataformas se limitara a medios periodísticos, expertos cualificados y datos contrastados. Luego puede haber las opiniones que se quiera. El debate sobre el cambio climático es un buen ejemplo, ya que hay diversidad de opiniones entre apocalípticos y negacionistas, pasando por alarmistas y condescendientes. No obstante, las opiniones deben estar fundamentadas sobre una base de información veraz que, puesta en contraste, ayude a crear opiniones fundadas.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los *Deepfake*s y por qué?

En la medida en que es algo peligroso hay que registrarlo, regularlo y combatirlo por el bien del individuo y de la sociedad. Además de las soluciones propuestas en la anterior pregunta, Enrique Campo afirma que los *Deepfakes* deben ser denunciados y combatidos a través de organismos especializados como el INCIBE, el CNI, los diferentes servicios contra el ciber-crimen de Policía y la Guardia Civil. Asimismo, también se debería contemplar la posibilidad de denunciarlos en los órganos dedicados a Consumo y Protección de los Consumidores. Contrariamente, Alberto Artero y Borja Adsuara creen que la respuesta no debería ir de la mano de los poderes públicos sino de agentes privados.

4.2.2. Cuestionario a expertos en política

En este cuestionario se ha contado con la participación de Carlos Fernández-Barbudo (Doctor en Ciencias Políticas y de la Administración y Relaciones Internacionales por la Universidad Complutense de Madrid), Javier Gil Pérez (profesor del Departamento de Relaciones Internacionales en la Universidad Pontificia Comillas Icade) y Carlos Rico (Doctor en Ciencia Política por la Universidad Autónoma de Madrid).

a) ¿Cree usted que los *Deepfakes* suponen una amenaza real para la democracia en el mundo?

Los tres entrevistados afirman que efectivamente, los *Deepfakes* suponen una amenaza real para la democracia en el mundo, que irá cobrando protagonismo en el futuro próximo. Carlos Rico explica que nuestros sistemas democráticos están basados en unas normas y procedimientos que se establecieron en un mundo que no es el actual. Las nuevas tecnologías han creado un universo de posibilidades, que los ingenieros que diseñaron los modelos institucionales y normativos de la democracia liberal nunca pudieron prever. Por lo tanto, hay un desfase entre las posibilidades reales actuales y los modelos institucionales y normativos de la democracia liberal. Para Carlos Fernández-Barbudo este desfase irá cobrando protagonismo en el futuro próximo, de tal modo que los *Deepfakes* acabarán por tomar un papel muy destacado dentro de la situación actual de desinformación producida por el auge de las redes sociales y la quiebra de confianza de los medios de comunicación de masas.

En relación con los riesgos a los que la sociedad y el sistema democrático se enfrentan, Carlos Rico observa que existe un problema a corto plazo y otro a largo plazo:

- **A corto plazo:** es difícil para el ciudadano distinguir lo que es verdad en las redes sociales e internet, y como es una tecnología que está muy establecida en nuestra realidad, se está creando un campo de cultivo sobre si es verdad lo que nos llega, ya que en ocasiones la información nos llega a través de gente de confianza y conocidos. El problema es que, si los *Deepfakes* están bien hechos, no son fáciles de distinguir y eso requiere al ciudadano un tiempo para identificarlo, tiempo que el ciudadano de a pie no tiene. La sociedad española está muy enfrentada, por eso corremos el riesgo de dar por buenos información y vídeos basura que nos hacen

tener una opinión muy negativa respecto al objetivo que ha sido manipulado: el gobierno, la oposición…

- **A largo plazo:** se está produciendo una devaluación del valor de la verdad (modelo de democracia liberal: nosotros elegimos a unos representantes y mediante el debate público, vemos si lo están haciendo bien y en las siguientes elecciones les juzgamos. Pero esto está basado en que en el debate público nosotros somos capaces de saber qué es lo que han hecho, cómo lo podrían haber hecho mejor…), pero si no sabemos lo que es verdad o mentira llegará un momento en que se desvirtúe el sistema democrático, porque dejamos de tener un espacio público compartido en el que todos podemos hablar de lo mismo.

Como afirmaba Alberto Artero, en un sistema democrático se confunde el tener derecho a expresar una opinión, con tener derecho a explicar diferentes hechos sobre una misma realidad. Carlos Rico afirma que no pueden existir diferentes hechos de una misma realidad, ya que eso niega la posibilidad de que exista cualquier tipo de debate. Esto es precisamente lo que hacen las *fake news* y de manera más compleja los *Deepfakes*: crear realidades diferentes para distintos tipos de personas, lo que provoca una devaluación de la verdad, una falta de credibilidad en las instituciones básicas de la democracia y al mismo tiempo una devaluación de la confianza en el parlamento, los partidos políticos, la oposición, las instituciones… Todo esto acabará provocando una enorme desafección por la democracia representativa.

b) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de *Deepfakes* sin condicionar la libertad de expresión?

Carlos Fernández-Barbudo y Javier Gil Pérez coinciden con los expertos en comunicación, en que prohibir el uso de *Deepfakes* no debería ser considerado una merma de la libertad de expresión, ya que esta libertad no ampara, en ninguna de sus formas, el derecho a mentir. Igualmente, creen que el debate no está tanto en la libertad de expresión y sus posibles limitaciones, como en las dificultades que tiene y tendrá el marco legal para la persecución de *Deepfakes*.

Por otro lado, Carlos Rico considera que lo que hace que este problema sea tan difícil de resolver es precisamente eso. La libertad de expresión, que es un derecho fundamental y

un principio básico de democracia liberal dice que todo el mundo tiene derecho a tener la opinión que quiera y a expresarla y difundirla. El problema es que no tenemos un "sistema quirúrgico" para poder distinguir en una opinión lanzada por cualquier medio, qué parte hay de hecho y qué parte hay de juicio de valor; normalmente lo mezclamos todo. El problema es que si interviene el Estado se generan miedos de que pueda limitar la libertad de expresión, pero si por otro lado el Estado deja libertad absoluta para que circule cualquier contenido, también puede ser muy peligroso para la capacidad de los ciudadanos para estar bien informados. Por ejemplo, si digo: "el gobierno es fascista" eso un juicio de valor, no un hecho. Tengo libertad de tener esa opinión, el problema está en que, en esa afirmación, si la comparto, puedo estar engañando a la persona que la recibe.

c) En el Estado de California ya se están aprobando leyes para prohibir y sancionar el uso de *Deepfakes* durante los meses de campaña electoral, ¿cree usted que España debería hacer lo mismo, o por lo contrario crear legislación con efecto permanente?

Para Carlos Rico y Carlos Fernández-Barbudo, el periodo de campaña electoral es un periodo en el que la libertad de información de calidad tiene que estar particularmente reforzado. En este sentido, como primer paso en la lucha contra la desinformación en sus diversas formas, se debería extender el control público del contenido que se publica en internet y las redes sociales. En cierta medida esto ya se hace: se garantizan ciertos debates para que los políticos puedan acceder a los medios públicos, se obliga a que los periodistas atiendan a políticos de todas las ideologías… el problema es que esto está pensado para los medios de comunicación tradicionales, y se tiene que adaptar también para redes sociales e internet. Carlos Fernández-Barbudo además añade, que esto debería aplicarse a las tecnologías de la desinformación en general, pero en su opinión los *Deepfakes* no pueden ser entendidos como parte de la libertad de expresión en el contexto de la lucha política más allá del uso humorístico que actualmente se hace de ellas.

En cambio, Javier Gil Pérez cree que el control de los *Deepfakes* debe ser integral, por tanto, no se debería ejercer únicamente durante los periodos de campaña electoral. El ser humano tiene un deseo natural de intentar controlarlo todo para sentirse más seguro, de ahí que hagamos leyes nacionales e internacionales, pero, aunque sobre el papel esto se prohíba, una cosa muy diferente es que se aplique y la manera en que se haga.

d) ¿Cree usted necesaria una regulación conjunta por parte de la Unión Europea para combatir este problema? ¿Y una herramienta de detección conjunta?

Todos coinciden en que el problema de la *post-verdad* y las *fake news* en todas sus variantes, no son un tema menor ni de los Estados por separado, sino que son un riesgo claro para la sostenibilidad y viabilidad de la democracia real como la conocemos. Para Carlos Fernández-Barbudo y Javier Gil Pérez, el principal reto es establecer técnicas estandarizadas y de confianza a nivel comunitario para identificar los *Deepfakes*. Una vez tengamos estos dispositivos técnicos no debería ser demasiado difícil hacer encajar su eliminación en el actual ordenamiento jurídico.

Para Carlos Rico, este problema también se debe abordar a nivel europeo, pero de manera simultánea en cuanto a regulación, educación ciudadana y avance técnico. Si la Unión Europea tiene en sus derechos fundamentales la defensa de la democracia, tiene que poner medidas a nivel europeo sobre la calidad del debate público en la democracia: preocupación ciudadana, cómo los periodistas tratan lo que hay en las redes, abordar soluciones técnicas de ciberseguridad, cómo mejorar nuestras capaces técnicas en la lucha contra el ciberdelito e identificar las plataformas donde se alojan la mayor parte de estos contenidos para eliminar las manipulaciones.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los *Deepfakes* y por qué?

Para Carlos Fernández-Barbudo, la respuesta que el Estado, no tanto el gobierno, debe dar ante este fenómeno es la que se recoge en la Estrategia de Seguridad Nacional, siempre en coordinación con otras agencias europeas para facilitar elementos de detección de los mismos y alerta temprana ante su divulgación.

Carlos Rico en cambio considera que, aunque probablemente no exista una respuesta tecnológica que evite esto, las grandes operadoras de internet deberían incorporar algunos sistemas para poder marcar o indicar al usuario que esa información puede ser falsa. Unido a su respuesta sobre el problema que supone a corto y largo plazo para la sociedad,

afirma que, a largo plazo se puede educar a los ciudadanos para que sean más críticos con ellos mismos. El ser humano es por naturaleza sectario y ve la política a través de la emoción en vez de la razón y distancia, por eso si no recibe educación crítica para darse cuenta de que es en efecto sectario, va a poder encontrar rápidamente información que reafirme sus creencias.

4.2.3. Cuestionario a expertos en ciberseguridad

Por último, en este cuestionario se ha contado con la participación de Ricardo Kleinlein (actualmente cursando un doctorado en Inteligencia Artificial por la Universidad Politécnica de Madrid), Guillem Colom (codirector de THIBER) y José María Blanco (Director de Inteligencia de Ciberseguridad en CIPHER).

a) ¿Cree usted que los expertos en ciberseguridad deberían contribuir a la educación del ciudadano en tener un mayor espíritu crítico respecto a las fake news y en concreto los Deepfakes? Sí la respuesta es que sí, ¿de qué manera deberían contribuir?

Según Ricardo Kleinlein, se ha fomentado una visión de la tecnología informática como algo oscuro e inaccesible. Tanto, que la sociedad ha asumido que funciona bien y es de confianza aún sin conocer su propiedad y funcionamiento. En resumen, la población no tiene, ni se ha preocupado por tener una visión crítica de estas tecnologías, así como de sus usos. Comenta, que dado el volumen de iniciativas de divulgación que existen, la comunidad investigadora ya está realizando su labor, al menos en gran medida. Él considera que la labor educativa no debe ir tanto de la mano de las empresas de ciberseguridad, sino que debe ser responsabilidad del propio ciudadano desmitificar estas tecnologías y ser crítico con ellas.

Por otro lado, tanto José María Blanco como Guillem Colom opinan que es necesario que todos los agentes de diferentes ámbitos contribuyan a que se mejore la concienciación sobre estos temas. Para Guillem Colom, además existe un elemento que es aún más preocupante en algunos casos, que es tratar de influenciar al legislador para que tome conciencia de ciertos problemas. La realidad es que las normas siempre van por detrás de la realidad, es decir, la regulación siempre se da a posteriori de que ocurra algo. Al final, lo que pasa con estas herramientas de innovación tecnológica, es que el legislador no las

conoce en profundidad, con lo cual se crea regulación tarde y mal. Por tanto, no es solo un tema de concienciación popular, sino también un tema político y regulador.

b) ¿Qué capacidad tienen las empresas de ciberseguridad en España para detectar *Deepfakes* y prevenir su propagación?

Ricardo Kleinlein afirma que aún estamos en una fase de desarrollo de "fakes" visuales creíbles y de perfeccionamiento, pero queda mucho que hacer. En este sentido, existen herramientas fácilmente accesibles cuya tasa de detección de *Deepfakes* es casi perfecta. El problema entonces, no es el escenario en que se encuentran los *Deepfakes* ahora, sino en el que pueden encontrarse en un futuro cuando se desarrollen herramientas hiperrealistas. Para José María Blanco, se deben considerar dos límites en la propagación de los *Deepfakes*: el respeto a los derechos humanos y la inducción al odio. En este sentido, en cuanto a la actuación de las empresas de ciberseguridad a este respecto, entra en juego la capacidad que tengan, así como su voluntad para hacerlo. Guillem Colom coincide con José María a este respecto, puesto que existen medios de comunicación y otras plataformas cuyo modelo de negocio se basa en la manipulación y la mentira.

Es cierto que a medida que esta tecnología evolucione y se perfeccione, también se plantea que los desarrollos en materia de inteligencia artificial puedan también mecanizar o facilitar significativamente la detección de *Deepfakes*. Sin embargo, Colom opina que el problema básico no es tecnológico, sino reputacional, ya que una vez se publica una información falsa pero llamativa, a pesar de los esfuerzos por desmentirla, la gente se queda con la primera noticia y en especial si es negativa.

c) ¿Cree necesaria la colaboración conjunta en la lucha contra los *Deepfakes*, entre medios de comunicación, redes sociales y otros agentes implicados con empresas de ciberseguridad?

Para los tres entrevistados, sí es necesaria la colaboración entre los agentes implicados. Sin embargo, Ricardo Kleinlein pone el énfasis en quién debería hacer qué. Para él, esto es algo más cuestionable, puesto que se deben evaluar primero los intereses de las partes. En este sentido, es imposible una lucha ideológica neutral contra los "fakes" en todas sus formas. Por otro lado, José María Blanco opina que actualmente existen muchos recursos para luchar contra esta amenaza.

Guillem Colom explica que, puesto en su conjunto, la desinformación es uno de los mayores problemas a nivel histórico que ha existido. Durante la Guerra Fría, la Unión Soviética y sus Estados Satélite eran expertos en realizar las llamadas "medidas activas" de comunicación, decían que era muy difícil meter la desinformación en países occidentales, por los estándares periodísticos de esos países (estaban bien pagados y verificaban las fuentes de las que venía la información). Sin embargo, ahora, dado que todo el mundo tiene acceso a plataformas de información, principalmente las redes sociales, este problema resulta mucho más difícil de acotar, puesto que verificar las fuentes de cada mensaje o artículo publicado puede ser una labor interminable. A parte de las redes sociales, existen muchos medios que buscan inmediatez y tampoco se preocupan por verificar las fuentes. Con lo cual, esto con los *Deepfake*s pasará cada vez más, porque se han convertido en un vector de propaganda política y, en definitiva, en una herramienta para perjudicar a los demás. Uno de los principales problemas, es que los mismos medios de comunicación son el problema. Dado que, para ciertos medios de comunicación publicar noticias intencionadamente falsas pero llamativas es parte de su modelo de negocio. En este sentido, todos los agentes implicados deberían coordinarse para poder enfrentarse a la post-verdad.

d) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que se debe dar para combatir la amenaza de los *Deepfakes* y por qué?

Mientras que todos coinciden en que la respuesta debe venir de la mano de todos los agentes implicados, Ricardo Kleinlein afirma que la respuesta institucional debe basarse en criterios meramente matemático-computacionales y su adopción debe ir acompañada de la crítica formada de los usuarios: cualquier otra respuesta estará dando un paso más hacia la censura. En cuanto a la respuesta tecnológica, mucha gente formada en este asunto habla de "implementar ética en los algoritmos", lo que considera que no es más que una medida de marketing disfrazada de innovación y buenas intenciones.

Para José María Blanco, los *Deepfakes* han llegado para quedarse, con lo cual lo más importante es buscar medidas de continuidad. Por tanto, la solución debe venir de la mano de un conjunto de medidas:

- Castigo ejemplar en los casos de faltas y delitos

- Educación y alfabetización digital, orientada a niños y jóvenes, de forma transversal a todas las asignaturas, y basada en el pensamiento crítico
- Autorregulación por parte de los propios medios de comunicación y la sociedad
- La colaboración de las empresas y operadores como Facebook o Twitter, entre otros
- Regulación institucional, definiendo mejor el tipo de falta o delito
- Técnicas de prevención y detección tecnológicas, mediante algoritmos para detectar fakes, blockchain...

Por otro lado, para Guillem Colom, lo óptimo sería que hubiera primero compromisos tácitos o manifiestos, de que todas las organizaciones políticas y sociales, al menos a nivel doméstico, intentaran evitar este tipo de cosas; sin embargo, viendo el ritmo de los tiempos esto es prácticamente imposible. Dentro de los países occidentales, a nivel electoral se está viendo cada vez más polarización en el voto, por lo tanto, cuando estos agentes (partidos políticos, medios de comunicación…) utilizan no solo la desinformación sino también los *Deepfake*, otra alternativa es que hubiera actores externos que trataran de influenciar sobre la política interna de los países, pero uno de los vectores más o igual de importantes, es en la misma política interna, la cantidad de herramientas de manipulación que se están utilizando para intentar ganar en las urnas. Por lo tanto, sí que sería necesaria una respuesta conjunta, pero al final, en muchos casos al final la propaganda política, para lo único que serviría, sería para ratificar ciertas preconcepciones que se tiene, por tanto, siempre va a haber intereses propios de un lado o de otro. Al final, nuestras fuentes de información nos hacen seleccionar de forma implícita las fuentes que más sirven para apoyar nuestros condicionantes sociales e incluso políticos, con lo cual cada vez estamos más metidos en una burbuja virtual.

V. CONCLUSIONES Y PROPUESTAS

El objetivo de este trabajo ha sido averiguar, en la medida de lo posible, si España está preparada o no para hacer frente a la amenaza de los *Deepfakes*. Por un lado, se ha hecho una revisión de la literatura para identificar los usos que se han hecho de esta tecnología junto con la respuesta que se ha dado hasta ahora por parte de la comunidad internacional. Una vez hecha la revisión de la literatura, se ha analizado la Estrategia de Ciberseguridad de España publicada en 2019, junto con la opinión de 9 expertos en comunicación, política y ciberseguridad. Para esta fase del trabajo, se han extraído las siguientes conclusiones:

- Dado que los estados miembros de la Unión Europea se decantan por un rol más paternalista, es lógico intuir que el gobierno español sigue también esta línea de pensamiento. Como se ha podido observar en la Estrategia Nacional de Ciberseguridad 2019, efectivamente este es el caso. Tras la publicación del Real Decreto Ley de seguridad de las redes y sistemas de información, el plan de acción del gobierno español está orientado a lograr un marco legal consolidado que penalice los ciber-ataques y garantice la seguridad, a través de un sistema policial y judicial más fuerte que defienda tanto al sector privado como a los ciudadanos españoles. El gobierno español apuesta por lograr autonomía tecnológica nacional mediante una mayor inversión en Innovación y Desarrollo, sin dejar de lado la cooperación con la Unión Europea en el fortalecimiento de los mecanismos de prevención de amenazas como los *Deepfakes*.

- En cuanto a los principales riesgos a los que se enfrenta la sociedad española, los expertos señalan que los *Deepfakes* pueden utilizarse para campañas de desprestigio y bulos, lo que a corto plazo generará una mayor desconfianza hacia los medios de comunicación masificados. El problema es que existe un desfase entre las posibilidades reales de manipulación y los modelos institucionales de la democracia. Puesto que el ser humano es sectario y necesita sentirse parte de un grupo para coexistir en la sociedad, los *Deepfakes* pueden generar un mayor enfrentamiento entre ciudadanos de diferentes ideologías, lo que a la larga hará que desaparezca todo tipo de debate político y que aumenten los extremismos, en

última instancia generando una gran desafección por el sistema democrático representativo como lo conocemos.

- Las plataformas digitales de comunicación deben trabajar proactivamente en la prevención de la difusión de información manipulada a través de imágenes. En este sentido, se propone el empleo de redes de *bots* para identificar patrones de comportamiento de estas tecnologías, así como la limitación de publicaciones que permitan hacer uso de esta tecnología de manera rápida y sencilla. Sin embargo, dado que en las redes sociales hay mayor libertad para publicar cualquier tipo de información, existe una frontera entre el creador de contenido y el distribuidor del mismo que hace que la responsabilidad que se puede exigir a las plataformas digitales sea limitada.

- Para evitar la desafección por la democracia representativa como la conocemos, los expertos en política reconocen que se debe hacer un esfuerzo por garantizar información de calidad como mínimo en los periodos de campaña electoral. En este sentido, hay diferentes opiniones sobre si esto debería hacerse únicamente de manera temporal o extendida en el tiempo. Además de esto, las *fake news* en todas sus variantes, no son un tema menor ni de los Estados por separado, sino que son un riesgo claro para la sostenibilidad y viabilidad de la democracia real como la conocemos. Por esta razón, se deben establecer en primer lugar técnicas estandarizadas y comunes a los estados miembros de la Unión Europea, además de hacer hincapié en la educación ciudadana e implementar un marco legal común que persiga este tipo de ciber-delitos.

- En el plano de la ciberseguridad, es crítica la colaboración conjunta entre medios de comunicación y esferas políticas para combatir la desinformación y educar al ciudadano para que tenga una actitud más crítica a la hora de procesar la información que recibe. El alcance que tienen las empresas de ciberseguridad para prevenir la propagación de *Deepfakes* cada vez será mayor a medida que esta tecnología evolucione, pero lo complicado ahora es enfrentarse a empresas cuyo modelo de negocio se sustenta en la publicación de contenido falso.

- En conjunto, las propuestas que se plantean para responder a esta amenaza son de tipo legal, técnico y educativo. El derecho a la libertad de expresión está regulado y tiene límites, con lo cual se debería perseguir el uso malicioso y la difusión de *Deepfakes* bajo el Código Penal español, además de permitir la posibilidad de denunciarlo a organismos como el CNI, INCIBE o la Guardia Civil. Complementariamente, lo óptimo sería dotar de cualidades periodísticas a la información que se publica, a través de autores y fuentes reconocidas que tengan responsabilidad económica en caso de cometer fraude. Para la detección temprana de este tipo de vídeos, el Estado español debería colaborar con las agencias europeas y con las grandes operadoras de internet, para indicar de alguna manera que la información que se está viendo puede ser falsa. Por último, es importante recalcar que los expertos en ciberseguridad afirman que la respuesta institucional debe darse únicamente mediante criterios matemático-computacionales, ya que cualquier otro tipo de respuesta sería dar un paso más hacia la censura.

A nivel personal, opino que lo más importante es recuperar la confianza de los ciudadanos en el gobierno español en sentido atemporal. Como se ha podido observar, el nivel de desconfianza es muy alto. Esto se podría conseguir a través de campañas de concienciación, leyes de transparencia junto con leyes que prohíban la desinformación. Una vez se vaya consiguiendo mayor confianza, lo que requerirá mucho tiempo, se deberán empezar a aplicar herramientas de verificación de la información. España debe invertir en herramientas de investigación y verificación de información para ganar la carrera entre "detectar un *Deepfake* y crearlo", puesto que ahora mismo se tarda más en lo primero que en lo segundo. El Estado, independientemente de las legislaturas políticas, debería sistemáticamente hacer responsables a las compañías de redes sociales, para que ellas también contribuyan a la detección y prevención de que los *Deepfakes* se divulguen.

VI. BIBLIOGRAFÍA

Administración General del Estado. (2019). *Estrategia Nacional de Ciberseguridad*. Gobierno de España: Ministerio de Presidencia, Relación con las Cortes e Igualdad.

Alto Analytics. (2017). The construction of anti-immigration electoral messages in Italy. *Alto Analytics*.

Andrés, R. R. (2017). Fundamentos del concepto de desinformación como práctica manipuladora en la comunicación política y las relaciones internacionales. *Ediciones Complutense*.

Bertolin, G. (2017). Digital hydra: security implications of false information online. *NATO StratcOM COE*.

Blanco, J. M. (2018). Verdad y mentira de la posverdad (II). *False Friends*.

Bobby Chesney, D. C. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*.

Bressan, S. (19 de September de 2019). Can the EU prevent deepfakes from threatening peace? *EU-Listco*.

Casanova, N. (2018). *Tecnologías multimedia empleadas como medio para desinformar*. Facultad de Ingeniería. Bogotá: Universidad Militar Nueva Granada.

Centro Criptológico Nacional. (2019). *Desinformación en el ciberespacio.* Ministerio de Defensa.

Coats, D. R. (2019). *Worldwide threat assessment of the US Intelligence Community*. Senate Select Committee on Intelligence.

Colomina, C. (2019). La desinformación de nueva generación: cinco escenarios políticos y geoestratégicos ante el fake. *Anuario Internacional CIDOB*.

Consejo de Seguridad Nacional. (2019). *Agencia Estatal Boletín Oficial del Estado*. Obtenido de https://www.boe.es/eli/es/o/2019/04/26/pci487/con

Cristian Vaccari, A. C. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, 1–13.

Darya Bazarkina, Y. P. (2019). Artificial Intelligence and new threats to International Psychological Security. *Russia in Global Affairs, 17*(1).

Edelman. (2018). *Edelman Trust Barometer. Global Report.*

El Diario. (2020). La Fiscalía identifica una decena de delitos en los que se puede incurrir por la difusión de bulos. *El Diario*.

EU-Listco. (25 de March de 2019). Anticipating governance breakdown and violent conflict. *EU-Listco*, 1.

European Commission. (2018). *Action plan against disinformation.*

European Commission. (2018). *Fake news and online disinformation.*

European Parliament. (2019). *Disinformation and propaganda - Impact on the functioning of the rule of law in the EU and its member states.* Policy Department for Citizen's rights and constitutional affairs.

European Union. (2015). *EU vs Disinfo: don't be deceived, question even more*.

Fernández-García, N. (2017). Fake news: una oportunidad para la alfabetización mediática. *Nueva Sociedad*(269), 1-2.

Galston, W. A. (8 de January de 2020). Is seeing still believing? The deepfake challenge to truth in politics. *Brookings*.

Garth Jowett, V. O. (2006). Propaganda and persuasion. *Sage*.

Gobierno de Aragón. (23 de Febrero de 2019). Primera guía del CNI contra las "fake news": contrasta fuentes, sospecha de anónimos y no compartas información dudosa. *Europa Press*.

Habermas, J. (2006). Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research1. *International Communication Association, 16*, 411–426.

Harwell, D. (2019). Top AI researchers race to detect 'deepfake' videos: 'We are outgunned. *The Washington Post*.

KAHN, J. (2020). The problem with the EU's A.I. strategy. *Fortune*.

Knight, W. (21 de June de 2019). A new deepfake detection tool should keep world leaders safe- for now. *MIT Technology Review*.

Kumar, S. (18 de Noviembre de 2019). Desinformación, deepfakes y democracia: un diálogo con Craig Silverman. *Ijnet: Red internacional de periodistas*.

Maldonado, M. A. (2016). La digitalización de la conversación pública: redes sociales, afectividad política y democracia. *Revista de estudios políticos*(173), 27-54.

Manuel Torres, C. E. (2019). *Desinformación: Poder y Manipulación en la era Digital.* Granada: Editorial Comares S.L.

Metz, C. (2019). Internet Companies Prepare to Fight the 'Deepfake' Future. *The New York Times*.

Meyer, R. (2018). The Grim Conclusions of the Largest-Ever Study of Fake News. *The Atlantic*, 1.

Ohm, P. (2010). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Review.*

RT News. (2017). Spain ‘won’t have enough tanks’: Catalonia to vote on independence, defy Madrid. *RT News*.

Sánchez, C. (3 de Diciembre de 2019). Deepfaking: no te fíes de todo lo que ves. *Blog Cuatrecasas*.

Schroeder, J. (2020). Free expression rationales and the problem of deepfakes within the EU and US Legal systems. *Southern Methodist University*, 6-7.

Stockler, A. (2019). MIT Deepfake Video 'Nixon Announcing Apollo 11 Disaster' Shows the Power of Disinformation. *Newsweek*.

Universidad de Navarra/ Oxforx University. (2018). Una audiencia diversa y preocupada por la desinformación. *Center for Internet studies and digital life*.

Vallés, J. M. (2010). Política democrática y comunicación: un rapto consentido. *Revista de estudios políticos*(150), 11-50.

Venkatararamakrishnan, S. (2019). Can you believe your eyes? How deepfakes are coming for politics. *Financial Times*.

Villasenor, J. (14 de February de 2019). Artificial Intelligence, Deepfakes and the uncertain future of truth. *Brookings*, 6.

VII. ANEXOS

CUESTIONARIO A EXPERTOS EN COMUNICACIÓN

Anexo 1. Cuestionario a Alberto Artero

a) ¿Cuál es el riesgo que asume la sociedad ante el fenómeno de los *Deepfake*?

Es más importante la causa que la consecuencia. Sólo analizando la primera nos permitirá contestar la segunda, esto es: los riesgos. Lo que busca el deepfake es la manipulación de las opiniones y/o percepciones de una realidad dotando de apariencia de realidad a lo que no son sino bulos. Cualquier iniciativa de este tipo tiene tres objetivos principales: uno social, de movilización de masas y creación de estados de opinión favorables o contrarios a algo; otro de poder, básicamente de ocupación de espacios de otros o nuevos; y, como colofón del anterior, económico. **El riesgo de la sociedad es la desinformación, la toma de decisiones inadecuadas**, por una parte; y el **conductismo, el aborregamiento social que lleva a aceptar cualquier realidad,** por otro. El ejemplo más claro de esto es lo que hacía Facebook con Cambridge Analytics para las campañas electorales: sin meter bulos alimentaban el rechazo a una opción política, alimentando otra. Solo se puede compensar con educación, pero da la impresión que esa es una batalla perdida.

b) ¿Por qué compensan las noticias falsas a quien se dedica a difundirlas?

Contestado previamente.

c) ¿Cuál es el papel que deben tener las plataformas digitales en las que se divulga información respecto a la lucha contra los *Deepfakes*?

Bueno, este es un punto muy interesante. Deben hacerlo estableciendo mecanismos de análisis y control que permita discriminar realidad y ficción. Sin embargo, no siempre es fácil: primero porque muchas veces no es fácil distinguirlas de partida, por eso es deepfake; en segundo término porque los recursos necesarios para hacerlo con precisión y rapidez serían tan ingentes que no compensarían financieramente; y tercero porque a veces hay profecías que se autocumplen: mentiras aparentes que devienen en verdad (imaginemos el tema del origen del Coronavirus en un laboratorio en Wuhan). Hay un punto aquí muy interesante y es que las redes sociales NO son medios de comunicación. En teoría cualquiera puede colgar lo que quiera y si luego es falso ampararse diciendo que era una recreación, una broma, un 'la guerra de los mundos' de Orson Wells. Y no hay responsabilidad. Otra cosa es que se delinca, esto es: que se suplanten identidades o

se manipulen informes. Pero esa frontera entre creador de contenido y distribuidor del mismo, es lo que hace que la responsabilidad que se puede exigir a estas plataformas sea limitada.

d) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de *Deepfakes* sin condicionar la libertad de expresión?

Como he dicho antes es muy complicado. La única manera que se me ocurre es dotar de cualidades periodísticas a las informaciones de modo tal que haya una identificación de fuentes solventes, un autor reconocido que responde de la veracidad de forma subyacente a un "nickname" y una responsabilidad económica en caso de irresponsabilidad. Lo idóneo sería que la distribución de contenidos por parte de las plataformas se limitara a medios periodísticos, expertos cualificados y datos contrastados. Luego puede haber las opiniones que se quiera (el cambio climático es un buen ejemplo con el debate que hay entre apocalípticos y negacionistas, pasando por alarmistas y condescendientes) pero es algo que tiene base y que, puesto en contraste, ayuda a crear opiniones fundadas.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los *Deepfakes* y por qué?

De este gobierno cabe esperar bien poco pero no es algo que sea aplicable a uno u otro gobierno. Las directrices generales son las que he subrayado en la pregunta anterior y que deberían ser de aplicación transversal a todas las democracias.

Anexo 2. Cuestionario a Enrique Campo

a) ¿Cuál es el riesgo que asume la sociedad ante el fenómeno de los Deepfake?

Ninguno. Para asumir un riesgo tienes que conocer la amenaza, y la sociedad no conoce ni valora el fenómeno de los Deepfake. Sí lo conocen los expertos en tecnología y comunicaciones, pero tampoco ellos son capaces de saber hasta dónde puede llegar el uso de estas técnicas y sus consecuencias.

Es como el coronavirus, si me permites la comparación. Unos pocos empezaron a hablar de él cuando surgió en Wuhan, China. Luego el virus comenzó a cobrarse vidas en ese país y se convirtió en noticia. Los medios de comunicación lo contamos y entrevistamos a epidemiólogos y otros expertos que tampoco fueron capaces de prever la tragedia que se avecinaba. El caso más emblemático es el de Fernando Simón, hablando de que en España quizás se daría algún caso aislado, pero nada preocupante. Y ya van 28.000 muertos. Estoy seguro de que, si preguntas por los deepfakes en cualquier redacción de cualquier medio de comunicación generalista español, 8 de cada 10 periodistas ni siquiera sabrán de qué les hablas.

La sociedad no asume ningún riesgo porque no conoce la amenaza. ¿Cuál es el riesgo real que afronta la sociedad? El de que el fenómeno se perfeccione, se generalice y llegue un momento en el que sea complicado distinguir un vídeo o audio manipulado de uno auténtico. Si eso ocurre, los medios de comunicación tendremos que ser aún más escrupulosos en la verificación de noticias. Sin embargo, lo que ocurra y corra como la pólvora por WhatsApp o Telegram será difícilmente desactivable.

b) ¿Por qué compensan las noticias falsas a quien se dedica a difundirlas? Pueden llegar a compensar por cuatro motivos:

- Por desinformar, difundiendo falsedades y haciendo ruido mediático en las redes sociales, para crear incertidumbre, inseguridad general o un estado de ánimo y de opinión determinado. Caso de quienes difunden noticias falsas o rumores interesados para desestabilizar relaciones entre potencias o entre gobernantes. Aún queda mucho por saber acerca de las injerencias rusas en procesos electorales europeos o norteamericanos.

- Por desprestigiar a alguien (persona, colectivo, empresa). Calumnia, que algo queda. Encender la mecha es muy fácil en Internet; apagarla, complicado.
- Por hacer negocio obteniendo visitas a una web o perfil social, logrando notoriedad gracias al tráfico redirigido a partir de la difusión de una noticia falsa. O mediante las típicas cadenas de mensajes que instan a reenviar, a través de las cuales también puedo obtener datos personales de usuarios e ir acumulándolos para luego venderlos.

  Otra forma de hacer negocio a través de las fake news es directamente la suplantación de identidades para engañar y obtener datos o dinero. Es el caso de quienes imitan a bancos o a organismos oficiales, haciéndose pasar por ellos y pidiendo por email datos personales o, directamente, transferencias económicas.
- Por diversión. Difundo noticias falsas por enredar, entretenerme, ver qué pasa, plantear un desafío, ver hasta dónde llegó...

c) ¿Cuál es el papel que deben tener las plataformas digitales en las que se divulga información respecto a la lucha contra los Deepfakes?

Deben actuar activamente contra ellas. Una cosa es que YouTube contenga vídeos que explican el proceso por el que se crea un deepfake y otra que permita tutoriales enseñando cómo hacerlos. Por otro lado, al igual que hay una tecnología para crearlos, pronto habrá una tecnología para detectarlos, filtrarlos e impedir su publicación o presencia en las plataformas digitales. Los "malos" siempre van por delante, pero la policía acaba pillándolos. Los operadores y plataformas son responsables de evitar la propagación de deepfakes y por tanto deben rechazarlos y denunciarlos.

d) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de Deepfakes sin condicionar la libertad de expresión?

El derecho a la libertad de expresión no ampara la comisión de delitos o engaños. El derecho a la libertad de expresión no es un derecho absoluto, sino que está regulado y tiene límites. Mi derecho a la libertad de expresión no ampara que yo pueda ir diciendo por ahí que tú copiaste en todos tus exámenes de Segundo y que por eso aprobaste. La libertad de expresión no ampara la mentira, la calumnia, la injuria.

Es perfectamente posible establecer un marco legal que prohíba el uso malicioso de deepfakes. Algunos estados norteamericanos ya lo han hecho. China también lo ha hecho, aunque, efectivamente, su marco jurídico no es homologable al de las democracias modernas. El uso y difusión maliciosa de deepfakes podría ser incluido sin problemas en el Código Penal como ya lo están otras prácticas que se desarrollan en el terreno virtual de Internet, quizás no tanto por la práctica concreta sino por el delito que se comete a través de ella, previamente tipificado y regulado. Es decir, aun no incluyendo expresamente las deepfakes en el Código Penal, ya podrían ser perseguibles legalmente en los casos en los que se observara un delito de injurias, calumnia, contra el honor o contra la propia imagen, por ejemplo.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los Deepfake y por qué?

Denunciarla y combatirla a través de organismos especializados como el INCIBE, el CNI, servicios contra el cibercrimen de Policía y Guardia Civil. Contemplarla en los órganos dedicados a Consumo y protección de los consumidores. Incluirla en el grupo de prácticas ilegales a perseguir y castigar cuando se busca la comisión de un delito.
¿Por qué? Porque en la medida de que es algo peligroso hay que registrarlo, regularlo y combatirlo por el bien del individuo y de la sociedad.

Anexo 3. Cuestionario a Borja Adsuara

a) ¿Cuál es el riesgo que asume la sociedad ante el fenómeno de los Deepfake?

El aumento de los bulos de desprestigio, la posibilidad de falsificar pruebas en juicios, las dificultades para identificar a alguien por reconocimiento facial, etc.

b) ¿Por qué compensan las noticias falsas a quien se dedica a difundirlas?

Hay potencias que buscan desestabilizar países, inversores que quieren hacer tambalear una empresa cotizada y ciudadanos que disfrutan engañando a gente.

c) ¿Cuál es el papel que deben tener las plataformas digitales en las que se divulga información respecto a la lucha contra los Deepfakes?

Luchar contra la viralización artificial del contenido falso (mediante redes de bots), para que, a base de repetición, pueda parecer verosímil.

d) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de Deepfakes sin condicionar la libertad de expresión?

Ya existen delitos que se pueden aplicar (de injurias, contra la integridad moral, de odio, de desórdenes públicos…). El único límite de la Libertad de Expresión es la Ley.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los Deepfake y por qué?

El Gobierno no debe entrar a combatir las Deepfake, igual que tampoco las fake news; si son constitutivas de delito debe actuar la policía, la fiscalía y los jueces.

CUESTIONARIO A EXPERTOS EN POLÍTICA

Anexo 4. Cuestionario a Carlos Fernández-Barbudo

a) ¿Cree usted que los Deepfakes suponen una amenaza real para la democracia en el mundo?

Los Deepfakes son un fenómeno novedoso que irá cobrando protagonismo en el futuro próximo. Esto se producirá de tal modo que acabarán por tomar un papel muy destacado dentro de la situación actual de desinformación producida por el auge de las redes sociales y la quiebra de confianza de los medios de comunicación de masas. En este sentido, los *Deepfakes* como parte protagonista de la desinformación deben ser considerados una amenaza real para los regímenes representativos de raigambre liberal, también conocidos como democracias constitucionales.

b) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de *Deepfakes* sin condicionar la libertad de expresión?

Prohibir el uso de *Deepfakes* no debería ser considerado una merma de la libertad de expresión, ya que esta libertad no ampara, en ninguna de sus formas, el derecho a mentir. Mucho menos puede considerarse parte de la libertad de expresión la modificación deliberada de un recurso audiovisual para hacer que una persona diga algo que jamás había dicho.

c) En el Estado de California ya se están aprobando leyes para prohibir y sancionar el uso de Deepfakes durante los meses de campaña electoral, ¿cree usted que España debería hacer lo mismo, o por lo contrario crear legislación con efecto permanente?

Considero que durante el periodo electoral deberían establecerse estrategias claras y transparentes para minimizar el impacto de la desinformación en el electorado. Esto es una cuestión compleja ya que afecta a libertades fundamentales de gran importancia. Fíjese que me refiero a la desinformación, en sus diversas formas, y no solo a las *Deepfakes*. Reitero que los *Deepfakes* no pueden ser entendidos como parte de la libertad de expresión en el contexto de la lucha política (otra cosa sería el uso humorístico).

d) ¿Cree usted necesaria una regulación conjunta por parte de la Unión Europea para combatir este problema? ¿Y una herramienta de detección conjunta?

No creo que estemos ante un problema legal novedoso. El principal reto es establecer técnicas estandarizadas y de confianza a nivel comunitario para identificar las DF. Una vez tengamos estos dispositivos técnicos no debería ser demasiado difícil hacer encajar su eliminación en el actual ordenamiento jurídico.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los *Deepfakes* y por qué?

Como parte prominente de la desinformación, la respuesta que el Estado, no tanto el gobierno, debe dar ante este fenómeno es el que se recoge en la Estrategia de Seguridad Nacional, siempre en coordinación con otras agencias europeas para facilitar elementos de detección de los mismos y alerta temprana ante su divulgación.

Anexo 5. Cuestionario a Javier Gil Pérez

a) ¿Cree usted que los Deepfakes suponen una amenaza real para la democracia en el mundo?

Si, es indudable. sobre todo por la contaminación que producen sobre la información veraz y en segundo término porque pueden producir importantes alteraciones de voto en base a información falsa o con contenido falso. Y eso es muy grave.

b) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de Deepfakes sin condicionar la libertad de expresión?

Yo creo que el debate no está tanto en la libertad de expresión y sus posibles limitaciones, como en las dificultades que tiene y tendrá el marco legal para la persecución de Deepfakes. Lo veo cuasi imposible porque la generación de Deepfake se encuentra en tierra de nadie. Una cosa es que tengamos la sospecha clara de que un determinado país u organización esté detrás de una determinada información, pero probarlo es algo muy distinto. Y regular esa zona gris es, a mi parecer, muy complicado.

c) En el Estado de California ya se están aprobando leyes para prohibir y sancionar el uso de Deepfakes durante los meses de campaña electoral, ¿cree usted que España debería hacer lo mismo, o por lo contrario crear legislación con efecto permanente?

Yo creo que a lo que tenemos que ir es a castigarlo de una manera integral, por ello no debe ser temporal. Ahora bien, como te he expresado antes, sobre el papel podemos prohibir lo que queramos, pero su aplicación práctica es diferente. El ser humano tiene un deseo natural de intentar controlarlo todo para sentirse mas seguro, de ahí que hagamos leyes nacionales e internacionales, pero la realidad es otra.

d) ¿Cree usted necesaria una regulación conjunta por parte de la Unión Europea para combatir este problema? ¿Y una herramienta de detección conjunta?
Si, debe ser una estrategia conjunta, a poder ser, pero ya sabes lo que pienso. Creo que es mas interesante fortalecer primero las capacidades de detección y neutralización.

Anexo 6. Cuestionario a Carlos Rico

a) ¿Cree usted que los Deepfakes suponen una amenaza real para la democracia en el mundo?

Sí. Nuestros sistemas democráticos están basados en unas normas y procedimientos que se establecieron en un mundo que no es el actual. Las nuevas tecnologías han creado un universo de posibilidades, que los ingenieros que diseñaron los modelos institucionales y normativos de la democracia liberal nunca pudieron prever. Por lo tanto, hay un desfase entre las posibilidades reales actuales y los modelos institucionales y normativos de la democracia liberal.

En cuanto al riesgo que estas tecnologías suponen, existe un problema a corto plazo y otro a largo plazo:

- **A corto plazo:** es difícil para el ciudadano distinguir lo que es verdad en las redes sociales e internet, y como es una tecnología que está muy establecida en nuestra realidad, se está creando un campo de cultivo sobre si es verdad lo que nos llega, ya que en ocasiones la información nos llega a través de gente de confianza y conocidos. El problema es que, si los Deepfake están bien hechos, no son fáciles de distinguir y eso requiere al ciudadano un tiempo para identificarlo, tiempo que el ciudadano de a pie no tiene. La sociedad española está muy enfrentada, por eso corremos el riesgo de dar por buenos información y vídeos basura que nos hace tener una opinión muy negativa del target: el gobierno, la oposición…
- **A largo plazo:** se está produciendo una devaluación del valor de la verdad (modelo de democracia liberal: nosotros elegimos a unos representantes y mediante el debate público, vemos si lo están haciendo bien y en las siguientes elecciones les juzgamos. Pero esto está basado en que en el debate público nosotros somos capaces de saber qué es lo que han hecho, cómo lo podrían haber hecho mejor…), pero si no sabemos lo que es verdad o mentira llegará un momento en que se desvirtúe el sistema democrático, porque dejamos de tener un espacio público compartido en el que todos podemos hablar de lo mismo.

Algo que se dice mucho en una democracia es: "yo tengo derecho a tener mi propia opinión". Eso es verdad, pero no tenemos derecho a tener **diferentes hechos** sobre una misma cosa, porque eso supone negar la posibilidad de cualquier tipo de debate. Esto es

lo que están haciendo las fake news, crear realidades diferentes para distintos tipos de personas, lo que provoca una devaluación de la verdad, una falta de credibilidad en las instituciones básicas de la democracia y al mismo tiempo una devaluación de la confianza en el parlamento, los partidos políticos, la oposición, las instituciones… lo que acabará provocando una enorme desafección por la democracia representativa, con lo cual es un problema de enorme calado.

b) ¿Hasta qué punto es posible establecer un marco legal que prohíba el uso de Deepfakes sin condicionar la libertad de expresión?

Lo que hace que este problema sea tan difícil de resolver es precisamente eso. Las fake news ponen de relieve que la diferencia entre la libertad de expresión y el derecho a la información es muy difícil de distinguir. Una cosa son los **hechos** (todos tenemos derecho a conocerlos de manera objetiva) y otra las **opiniones** y los juicios de valor, que son libres.

La libertad de expresión, que es un derecho fundamental y un principio básico de democracia liberal dice que todo el mundo tiene derecho a tener la opinión que le de la gana y a expresarla y difundirla. El problema es que no tenemos un "sistema quirúrgico" para poder distinguir en una opinión lanzada por cualquier medio, qué parte hay de hecho y qué parte hay de juicio de valor, normalmente lo mezclamos todo. Cuando tengo la libertad de mandar un tweet o reenviar un contenido por Whatsapp etc, puedo estar mezclando tanto opiniones como hechos, y ¿quién es el juez que está legitimado para decidir qué parte hay de hechos y qué parte hay de juicios de valor? Es muy difícil, salvo que un ciudadano me denuncie sobre injurias o calumnias, un juez podrá juzgarme, pero no hay un sistema fuera de eso de limitar los tweets etc.

El problema es que si interviene el Estado se generan miedos de que pueda limitar la libertad de expresión, pero si por otro lado el Estado deja libertad absoluta para que circule cualquier contenido, también puede ser muy peligroso para la capacidad de los ciudadanos para estar bien informados. Por ejemplo si digo: "el gobierno es fascista" eso un juicio de valor, no un hecho. Tengo libertad de tener esa opinión, el problema está en que, en esa afirmación, si la comparto, puedo estar engañando al que la recibe.

c) En el Estado de California ya se están aprobando leyes para prohibir y sancionar el uso de Deepfakes durante los meses de campaña electoral, ¿cree usted que España debería hacer lo mismo, o por lo contrario crear legislación con efecto permanente?

La esfera pública debería tener un entorno público constituido por medios de comunicación tradicionales y debería tener una protección en la democracia, pero como eso es muy difícil, una medida correctora o paliativa, tendría que ser que en el periodo de campaña electoral, que es un periodo en el que la libertad de información de calidad tiene que estar reforzado, se extendiese el control público del contenido que se publica en internet y las redes sociales. En cierta medida esto ya se hace: se garantizan ciertos debates para que los políticos puedan acceder a los medios públicos, se obliga a que los periodistas atiendan a políticos de todas las ideologías… el problema es que esto está pensado para los medios de comunicación tradicionales, y se tiene que adaptar también para redes sociales e internet.

Como primer paso esto debería hacerse, porque sino agentes externos o grupos anti-sistema pueden utilizar información para desestabilizar el debate público en un momento muy delicado, que es cuando se determina quién va a gobernar. Debe ser un periodo especialmente protegido. Eso exige hacer reformas para controlar todos los medios de comunicación (tradicionales y nuevos).

d) ¿Cree usted necesaria una regulación conjunta por parte de la Unión Europea para combatir este problema? ¿Y una herramienta de detección conjunta?

La Unión Europea tiene que considerar el problema general de la posverdad, las fake news y todos sus derivados. Esto no es un tema menor ni de los Estados por separado, es un riesgo claro para la sostenibilidad y viabilidad de la democracia real como la conocemos. Si la Unión Europea tiene en sus derechos fundamentales la defensa de la democracia, tiene que poner medidas a nivel europeo sobre la calidad del debate público en la democracia: preocupación ciudadana, cómo los periodistas tratan lo que hay en las redes, abordar soluciones técnicas de ciberseguridad, cómo mejorar nuestras capaces técnicas en la lucha contra el ciberdelito e identificar las plataformas donde se alojan la mayor parte de estos contenidos para eliminar las manipulaciones.

Los Deepfakes forman parte de un contexto mucho más dañino. Los partidos políticos han apostado por la polarización, es decir, el enfrentamiento brutal contra tu rival (da muchos votos). Cuando los asesores de los partidos políticos tienen un material en internet que puede ser usado contra su rival, hay una tentación muy grande de usarlo y cuando un político de rango lo dice en un debate, le está dando un peso que no hubiera tenido nunca. Se mezclan políticos que buscan la polarización, con ciudadanos cada vez más sectarios, con medios de comunicación que buscan llamar la atención y con gente que tiene recursos tecnológicos que nunca antes se habían utilizado.

e) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno español para combatir la amenaza de los Deepfake y por qué?

Esto es un problema mucho más grave de lo que parece, porque ¿cómo podemos controlar lo que cada uno de los ciudadanos que vive en un país, en el ejercicio de su libertad, habla de manera veraz?

No creo que haya una respuesta tecnológica que evite esto, pero las grandes operadoras de internet (Facebook, Whatsapp, google…) deberían incorporar algunos sistemas para poder marcar o indicar al usuario que esa información puede ser falsa. Tienen que trabajar mejor para poder filtrar la información e informar al usuario.

A largo plazo se puede educar a los ciudadanos para que sean más críticos con ellos mismos. Nosotros estamos acelerando esto con nuestro sectarismo, y vemos la política con emoción y no con razón y distancia, somos los primeros agentes que estamos recibiendo mercancía averiada y creyéndonosla. Si yo no tengo una educación crítica en ser capaz de darme cuenta de que soy muy sectario, voy a poder encontrar muy fácilmente información que reafirme mi creencia.

Es curioso porque la manipulación, las visiones parciales, las exageraciones, el juego de lenguaje… lleva pasando décadas en la política, el problema es que con las redes sociales empezamos a ver que esto es insostenible. Si los políticos tienen tendencia a utilizar el lenguaje para sus intereses partidistas, los medios de comunicación lo hacen para tener más audiencia… Esta situación se convierte en un caldo de cultivo.

CUESTIONARIO A EXPERTOS EN CIBERSEGURIDAD

Anexo 7. Cuestionario a Ricardo Kleinlein

a) ¿Cree usted que los expertos en ciberseguridad deberían contribuir a la educación del ciudadano en tener un mayor espíritu crítico respecto a las fake news y en concreto los Deepfakes? Sí la respuesta es que sí, ¿de qué manera deberían contribuir?

Se ha fomentado una visión de la tecnología informática como algo oscuro e inaccesible. Tanto, que hemos asumido que funciona bien y es de confianza aún sin conocer su propiedad y funcionamiento (GPS vs Galileo, Android/iOS...). En resumen, la población no tiene, ni se ha preocupado por tener una visión crítica de estas tecnologías, así como de sus usos (noticias, software, riesgos...). Dado el volumen de iniciativas de divulgación que existen, la comunidad investigadora ya está realizando su labor, al menos en gran medida. El ciudadano medio más que formarse en estas tecnologías, necesita desmitificarlas y ser crítico con ellas (en el sentido de crítica constructiva).

b) ¿Qué capacidad tienen las empresas de ciberseguridad en España para detectar Deepfakes y prevenir su propagación?

Los Deepfakes a día de hoy, son relativamente fáciles de identificar y no suponen una gran amenaza. Si mi tesis llega a buen puerto puede que el escenario cambie, pero mientras tanto, generar rostros que de verdad sean creíbles por el usuario está complicado, y existen herramientas fácilmente accesibles cuya tasa de detección de fakes es casi perfecta. Aún estamos en una fase de desarrollo de fakes visuales creíbles y de perfeccionamiento, pero queda mucho que hacer. Otra cosa son las fake news...

c) ¿Cree necesaria la colaboración conjunta en la lucha contra los Deepfakes, entre medios de comunicación, redes sociales y otros agentes implicados con empresas de ciberseguridad?

La pregunta así redactada es, disculpa y lo digo sin intención de ofender, capciosa. Sí, por supuesto que hace falta trabajo multidisciplinar, pero quién debe hacer qué es más cuestionable. Eso implica primero evaluar los intereses de las partes. Como comprenderás, es imposible una lucha ideológicamente neutral frente a estos fakes, por

lo que el tema empieza a resultar interesante. ¿Ves por dónde voy? ¿Debemos llamarlo "lucha"? ¿Debemos esperar una respuesta legislativa, de control o simplemente desarrollar respuestas tecnológicas?

d) ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que se debe dar para combatir la amenaza de los *Deepfake* y por qué?

Por desgracia, incluso gente formada en este campo habla de cosas como "implementar ética en los algoritmos". Simple y llanamente dicho, eso es mero marketing vacuo disfrazado de innovación y buenas intenciones. Los detalles de por qué esto es así se escapan un poco de este mensaje. La respuesta debe venir de la mano de la desmitificación de estos algoritmos, y su adopción y crítica formada por parte de los usuarios.

Artificial viene del latín artis ("arte") y facere ("hacer"). El hecho de que a este campo se le llame Inteligencia Artificial, lleva a la gente a pensar que estas herramientas tienen un componente antropomórfico y de resolución de problemas, cuando son meros "destornilladores con botones". Cualquier respuesta institucional que busque la eliminación de contenido potencialmente falso y cuyo criterio no esté basado exclusivamente en criterios matemático-computacionales estará dando un paso más hacia la censura, siendo especialmente cierto en el caso de los *Deepfakes,* y moderadamente cierto en las fake news.

Anexo 8. Cuestionario a José María Blanco

1. ¿Cree usted que los expertos en ciberseguridad deberían contribuir a la educación del ciudadano en tener un mayor espíritu crítico respecto a las fake news y en concreto los Deepfakes? Sí la respuesta es que sí, ¿de qué manera deberían contribuir?

Sin duda, los Deepfakes ya son una amenaza en el presente. Hay diversos motivos, pero el principal es la forma en la que están contribuyendo a polarizar nuestras sociedades, lo que redunda en una pérdida de cohesión social. De manera directa están induciendo odio, racismo y produciendo radicalización. En este sentido, debería haber una colaboración conjunta por parte de todos los agentes implicados para la concienciación del ciudadano.

2. ¿Qué capacidad tienen las empresas de ciberseguridad en España para detectar *Deepfakes* y prevenir su propagación?

Por un lado, para que algo sea perseguible, en un estado democrático, es fundamental la tipicidad de la posible falta o delito. Evidentemente la mentira por sí misma no lo es, pero sí que puede caer en diferentes tipologías. A nivel personal creo que debemos considerar dos límites: el respeto a los derechos humanos y la inducción de odio. En cuanto a la actuación de las empresas de ciberseguridad a este respecto, entra en juego la capacidad que tengan, así como su voluntad para hacerlo.

3. ¿Cree necesaria la colaboración conjunta en la lucha contra los Deepfakes, entre medios de comunicación, redes sociales y otros agentes implicados con empresas de ciberseguridad?

Sí, es necesario desarrollar esfuerzos conjuntos, además también por parte de la Unión Europea. Pero si buscas la información se están haciendo bastantes cosas, y son ya muchos los recursos existentes.

4. ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que debe dar el gobierno Español para combatir la amenaza de los Deepfake y por qué?

Esto ha llegado para quedarse, por lo que hay que buscar medidas de continuidad. Llegó tal y como lo conocemos, porque ha existido siempre. Pero internet y las nuevas tecnologías son un facilitador y potenciador. Entiendo que la solución debe ser un conjunto de medidas:

- Castigo ejemplar en los casos de faltas y delitos
- Educación. Alfabetización digital
- Autorregulación. Los propios medios de comunicación y sociedad
- Las empresas, como los operadores (Facebook, Twitter...)
- Una parte puede ser legal, definiendo mejor el tipo de falta o delito. La alfabetización digital es fundamental, orientada a niños y jóvenes, de forma transversal a todas las asignaturas, y basado en el pensamiento crítico. Y sin duda, con tecnología: algoritmos para detectar fakes, blockchain...

Anexo 9. Cuestionario a Guillem Colom

1. ¿Cree usted que los expertos en ciberseguridad deberían contribuir a la educación del ciudadano en tener un mayor espíritu crítico respecto a las fake news y en concreto los Deepfakes? Sí la respuesta es que sí, ¿de qué manera deberían contribuir?

En la era de la desinformación, ha habido también una proliferación de expertos en este asunto, lo que se desconoce si hasta cierto punto esto está siendo más un problema que una solución. De todos modos, sí que creo que se debe contribuir a combatir esta amenaza, pero no ya solo los expertos en ciberseguridad: es necesario que todo el mundo de distintos ámbitos contribuya a que se mejore la concienciación sobre estos temas. Un *Deepfake* no deja de ser la aplicación de una tecnología para manipular la verdad, lo cual hasta cierto punto siempre se ha hecho, quizás a un nivel menos elaborado. Por ejemplo, durante la Guerra Fría en la Unión Soviética era común la práctica de borrar a personas controvertidas de fotografías políticas.

Además, existe un elemento que es aún más preocupante en algunos casos, que es tratar de influenciar al legislador para que tome conciencia de ciertos problemas. La realidad es que las normas siempre van por detrás de la realidad, es decir, la regulación siempre se da a posteriori de que ocurra algo. Al final, lo que pasa con estas herramientas de innovación tecnológica, es que el legislador no las conoce en profundidad, con lo cual se crea regulación tarde y mal. Por tanto, no es solo un tema de concienciación popular, sino también un tema político y regulador. Además, ahora las redes sociales se han convertido en una caja de resonancia en la que personas aparentemente anónimas, tienen cierta libertad para exponer la verdad que ellos desean que se vea. En el caso de las últimas elecciones de Brasil, lo que puso a Bolsonaro en el poder fue hasta cierto punto por la información que se difundía a través de Whatsapp. Los Deepfakes necesitan un medio de transmisión, el problema viene cuando estas herramientas llegan a los medios de comunicación, quienes los replican y los legitiman más.

2. ¿Qué capacidad tienen las empresas de ciberseguridad en España para detectar Deepfakes y prevenir su propagación?

En este sentido, entra en juego la capacidad de las empresas y su voluntad para hacerlo. Hace poco Salió una noticia de que Facebook no solo había hecho oídos sordos a muchas

cosas en la campaña política a favor de Trump, sino que a través de sus propios algoritmos, había apoyado de forma directa a la politización de las campañas de Trump, es decir, que había un cierto modelo de negocio al respecto. En este sentido, se ha demostrado que Facebook tenía capacidad para detectar posibles alteraciones digitales de los Deepfake más sencillos, que se publicaban en su plataforma. Es cierto que a medida que esta tecnología evolucione y se perfeccione, también se plantea que los desarrollos en materia de inteligencia artificial pueden también mecanizar o facilitar muchísimo la detección de Deepfakes. Sin embargo, desde mi punto de vista, el problema básico no es un problema tecnológico, sino reputacional, ya que una vez se publica una información falsa pero llamativa, a pesar de los esfuerzos por desmentirla, la gente se queda con la primera noticia y en especial si es negativa.

3. ¿Cree necesaria la colaboración conjunta en la lucha contra los Deepfakes, entre medios de comunicación, redes sociales y otros agentes implicados con empresas de ciberseguridad?

Sí, sería fundamental, lo que pasa es que puesto en su conjunto, uno de los máximos problemas a nivel histórico que se ha tenido con el tema de la desinformación, es que durante la Guerra Fría, la Unión Soviética y sus estados satélite eran expertos en realizar medidas activas (desinformación, propaganda negra…), decían que era muy difícil meter la desinformación en países occidentales, por los estándares periodísticos de esos países (estaban bien pagados y verificaban las fuentes de las que venía la información).

Ahora, dado que todo el mundo tiene acceso a plataformas de información, principalmente las redes sociales, este problema resulta mucho más difícil de acotar, puesto que verificar las fuentes de cada mensaje o artículo publicado puede ser una labor interminable. A parte de las redes sociales, existen muchos medios que buscan inmediatez y tampoco se preocupan por verificar las fuentes. Con lo cual, esto con los Deepfake pasará cada vez más, porque se han convertido en un vector de propaganda política y, en definitiva, para perjudicar a los demás. Uno de los problemas, es que los mismos medios de comunicación son el problema. Dado que, para ciertos medios de comunicación, publicar noticias intencionadamente falsas pero llamativas, es parte de su modelo de negocio. En este sentido, todos los agentes implicados deberían coordinarse para poder enfrentarse a la post-verdad. Soy especialmente pesimista a este respecto, porque hay muchos modelos de negocio fundamentados en mentir para jugar al despiste.

4. ¿Qué tipo de respuesta (tecnológica, social, política…) cree usted que se debe dar para combatir la amenaza de los *Deepfake* y por qué?

Lo óptimo sería que hubiera primero compromisos tácitos o manifiestos, de que todas las organizaciones políticas y sociales, al menos a nivel doméstico, intentaran evitar este tipo de cosas, sin embargo, viendo el ritmo de los tiempos esto es prácticamente imposible. Dentro de los países occidentales, a nivel electoral se está viendo cada vez más polarización en el voto, por lo tanto, cuando estos agentes (partidos políticos, medios de comunicación…) utilizan no solo la desinformación sino también los *Deepfake*, otra alternativa es que hubiera actores externos que trataran de influenciar sobre la política interna de los países, pero uno de los vectores más o igual de importantes, es en la misma política interna, la cantidad de herramientas de manipulación que se están utilizando para intentar ganar en las urnas. Por lo tanto, sí que sería necesaria una respuesta conjunta, pero al final, en muchos casos al final la propaganda política, para lo único que serviría, sería para ratificar ciertas preconcepciones que se tiene, por tanto, siempre va a haber intereses propios de un lado o de otro. Al final, nuestras fuentes de información, nos hacen seleccionar de forma implícita las fuentes que más sirven para apoyar nuestros condicionantes sociales e incluso políticos, con lo cual cada vez estamos más metidos en una burbuja virtual.