POR QUÉ LAS CAMPAÑAS ELECTORALES COGNITIVAS BASADAS EN LA POSVERDAD PUEDEN EROSIONAR LAS DEMOCRACIAS DE OPINIÓN. CÓMO AFECTA A ESTE TIPO DE CAMPAÑAS LAS MENTIRAS PROFUNDAS DE BASE AUDIOVISUAL BASADAS REDES NEURONALES GENERATIVAS

Por

LUIS MIGUEL GONZÁLEZ DE LA GARZA
Departamento de Derecho Constitucional UNED

Revistas@iustel.com

Revista General de Derecho Constitucional 32 (2020)

RESUMEN: Si podemos convenir en que cada persona puede tener su propia opinión sobre cualquier tema, nadie está en cambio autorizado a tener sus propios hechos sobre tales temas. Esto, que en el paradigma analógico era cierto puede dejar de serlo en el paradigma de la realidad virtual, singularmente cuando estos hechos pueden ser falsificados de forma tan precisa, eficaz y poco costosa que la realidad puede dejar de ser lo que hasta ahora ha sido para convertirse en un campo de excentricidades, dudas, omisiones de diseño intencional y certezas construidas ad hoc.

La protección de la verdad en la democracia del siglo XXI es la esencia que hay que preservar frente a los ataques que van a ir in crescendo, primero fueron las Fake News, ahora son las Deepfakes y veremos nuevas dimensiones de estas patologías informacionales, pero ambas tienen un denominador común, su ataque a la verdad mediante la tergiversación de los hechos. Siendo el objetivo de estas distorsiones de muy diversa naturaleza, los objetivos de manipulación electoral son los que en este trabajo nos van a ocupar ya que afectan a un elemento medular de la democracia de opinión, una opinión pública nutrida de afluentes no tóxicos que permitan a los ciudadanos formarse juicios veraces basados en hechos ciertos. Dado que en la política la verdad importa, al final podría importar más que cualquier otra cosa es esencial abordar este campo de estudio explicando la naturaleza de las Deepfakes y proponiendo algunas soluciones para su control jurídico.

PALABRAS CLAVE: Posverdad, redes neuronales, deepfakes, democracia, falsedad, veracidad, propaganda, autocensura, feudalismo.

WHY POST-TRUTH-BASED COGNITIVE ELECTORAL CAMPAIGNS CAN ERODE OPINION DEMOCRACIES. HOW DEEP LIES OF AUDIOVISUAL

BASE BASED ON GENERATIVE NEURAL NETWORKS AFFECT THIS TYPE OF CAMPAIGNS

ABSTRACT: If we can agree that each person can have their own opinion on any subject, nobody is instead authorized to have their own facts on such topics. This, which was true in the analogue paradigm, may cease to be true in the virtual reality paradigm, especially when these facts can be falsified so accurately, efficiently and inexpensively, that reality can cease to be what until now it has been to become a field of eccentricities, doubts, omissions of intentional design and certainties built ad hoc.

The protection of the truth in the democracy of the 21st century is the essence that must be preserved against the attacks that will go in crescendo, first it was the Fake News, now they are the Deepfakes and we will see new dimensions of these informational pathologies, but both they have a common denominator, their attack on the truth by misrepresenting the facts. Being the objective of these distortions of a very diverse nature, the objectives of electoral manipulation are those that in this work will occupy us since they affect a core element of the democracy of opinion, a public opinion nourished by non-toxic tributaries that allow Citizens form truthful judgments based on true facts. Given that in politics the truth matters, in the end it could matter more than anything else it is essential to address this field of study explaining the nature of the Deepfakes and proposing some solutions for their legal control.

KEY WORDS: Post-truth, neural networks, deepfakes, democracy, falsehood, veracity, propaganda, self-censorship, feudalism.

Fecha de recepción: 24/05/2019

Fecha de aceptación: 23/10/2019

1. CONSIDERACIONES PRELIMINARES

Sin libertad de pensamiento no puede existir cosa alguna como la sabiduría, y tampoco algo como la libertad pública sin libertad de expresión (Gordon: 1721)[1]. De alguna forma todo aquello que se dirija a la manipulación de la libertad de pensamiento falseando los hechos afecta directamente a la libertad de expresión, ya que expresar libremente un pensamiento manipulado o deliberadamente erróneo no es a la postre sino efecto de una causa a la que los poderes públicos deben prestar especial atención. No cabe duda de que la verdad es esencial para el desarrollo de una sociedad libre y justa, como señalara (Frankfurt: 2007)[2] las civilizaciones *nunca* han podido prosperar, ni podrán hacerlo, sin cantidades ingentes de *información fiable sobre los hechos*. Tampoco pueden florecer si están acosadas por las problemáticas infecciones de creencias *erróneas.* Para crear y mantener una cultura avanzada es preciso que no nos

[1] Gordon, Thomas, Carta 15. Sobre la libertad de expresión: que resulta inseparable de la libertad pública, en: Cueva Fernández, Ricardo, "*Catas de Catón*", BOE, Madrid, 2018, pág. 43.

[2] Frankfurt, Harry G, "*Sobre la charlatanería y sobre la verdad*", Paidos, Barcelona, 2007, pág. 82-83.

dejemos debilitar por el error y la ignorancia. Necesitamos saber un gran número de verdades, y también, desde luego como hacer un uso productivo de ellas. El problema medular de la información falsa es precisamente ese, es un ataque directo a la cognición humana que es la base del desarrollo individual y colectivo de la sociedad, no se trata de tener opiniones se trata de que la información falsa pretende crear hechos falsos sobre los que se edifican las opiniones.

Las noticias falsas, la desinformación y la posverdad tienen una larga tradición. El uso del fraude, la falsificación y otras formas de engaño para influir en la política (Chesney y Citron: 2019)[3] no es nada nuevo, por supuesto. Cuando el USS Maine explotó en el puerto de La Habana el 15 de febrero de 1898, los tabloides estadounidenses de William Randolph Hearst y Joseph Pulitzer utilizaron relatos incendiarios y falaces del incidente para incitar al público a la guerra con España.

Las noticias falsas de naturaleza virtual y de difusión masiva a través de las redes sociales -como uno de sus vectores de difusión- juntamente con el correo electrónico o las redes como Facebook, WhatsApp, Telegram, etc., fueron y son un hecho social relativamente reciente de un tipo de propaganda electoral que podría datarse su inicio para el gran público en las elecciones presidenciales de 2016 en los Estados Unidos. Se puede encontrar en la actualidad una amplia base de documentación científica sobre estas, así como el interés de la comunidad jurídica por ofrecer mecanismos de detección y mitigación de estas fuentes de información fraudulenta que tratan de generar diversos efectos cognitivos como, por ejemplo, la desinformación a través de desvirtuar la información veraz con información que no lo es o no lo es en diversos grados, formulas en suma de demoler las concepciones privadas y públicas generales sobre la certidumbre de los hechos. El engaño tiene como objetivo producir estrategias o adoptar decisiones subóptimas en el sistema cognitivo de la víctima. Es decir, el engaño altera las creencias de la víctima y tal circunstancia modifica las decisiones de esta.

Desde 2016 el fenómeno sigue en aumento como han estudiado (Marchal y colaboradores: 2018)[4] incrementándose el volumen de noticias falsas en las redes sociales y disminuyendo el uso de fuentes solventes a menos del 5 por ciento en tales redes sociales[5].

---

[3]Chesney Robert y Daniel Citron, “Deepfakes and the New Disinformation War. The Coming Age of Post-Truth Geopolitics”, *Foreign Affairs*, January/February, 2019, págs. 147-155.

[4] Marchal, Nahema, Lisa-Maria Neudert, Bence Kollanyi, Philip N. Howard, “Polarization, Partisanship and Junk News Consumption on Social Media During the 2018 US Midterm Elections” Comprop, Data Memo 2018.5, noviembre 1, 2018. https://comprop.oii.ox.ac.uk/research/midterms2018/ (última visualización, 7 de marzo de 2019)

[5] La cantidad de noticias falsas en circulación en las redes sociales es mayor que durante las elecciones presidenciales de los EE. UU., de 2016, con usuarios que comparten más noticias falsas que noticias profesionales en general, las noticias falsas una vez consumidas por la base de

Nuestro propósito en este trabajo es considerar una nueva dimensión de un viejo fenómeno -el uso de la mentira (Koyré: 2015)[6] con finalidad política- que puede situarse en la órbita de las noticias falsas o más bien de la información falsa pero cuya característica relevante no es su forma escrita, que podría ser el modelo típico de noticia falsa o *fake news*. Nos referimos a las denominadas "*Deepfakes*" o *mentiras profundas* expresadas a través de la síntesis de imágenes digitales que emulan situaciones reales. Estas pueden tener una efectividad o impacto en los procesos electorales de alto riesgo para lo que debería ser un proceso de formación de la opinión pública justo basado en la veracidad de la información que llega a los ciudadanos por los diversos canales de información disponibles.

La capacidad para inspirar miedo, temor, duda o desazón, seguramente son algunos de los efectos más perniciosos cuando se diseñan con esa finalidad específica ya que ese miedo se volverá viral con extrema facilidad. En el fondo, tales imágenes están dirigidas a las emociones y una difusión de esas imágenes sin claros límites jurídicos pueden erosionar gravemente la confianza de los ciudadanos en todos los medios, tanto en los veraces como en los que lo son menos o los que no lo son en absoluto. Las noticias falsas operan destruyendo la credibilidad de las narrativas veraces, cuestionándolas con cursos narrativos falaces que, por su proximidad con la realidad hacen dudar o incurrir en error a los públicos objetivos, entendiendo por tales aquellos a los que se les dirige un contenido específico porque previamente se ha estudiado su sensibilidad hacia el temor que se pretende explotar mediante las técnicas OCEAN de clasificación por perfiles psicométricos.

Cuando hablamos de públicos objetivos, con (Quattrociochi: 2018)[7] hacemos referencia a que los usuarios de las redes sociales están polarizados, se informan y forman su opinión según un proceso cognitivo o sesgo, bien conocido que evita el conflicto apoyando las narrativas que respaldan sus propias creencias. Los contenidos se seleccionan por *sesgo de confirmación*[8] eso conlleva la creación de grupos uniformes u homogéneos en torno a temas y discursos específicos los cuales tienden a reforzarse

---

apoyo del presidente Trump y la extrema derecha ahora está siendo consumida por más usuarios conservadores de las redes sociales, y menos del cinco por ciento de las fuentes informativas a las que se hace referencia en las redes sociales son de agencias públicas, expertos o los mismos candidatos políticos.

[6] Koyré, Alexander, "*La función política de la mentira moderna*", Pasos perdidos, Madrid, 2015.

[7] Quattrociocchi, Walter, "La era de la (desinformación)", *Investigación y Ciencia*, nº 91, 2018, págs.30-31.

[8] El sesgo de confirmación es la tendencia de una persona a favorecer y priorizar la información que confirma sus suposiciones, creencias previas, ideas preconcebidas o hipótesis, independientemente de que éstas sean verdaderas o falsas, tratando así de evitar la tensión emocional que produce un estado de disonancia cognitiva que es el sentimiento de malestar que se experimenta cuando una idea entra en conflicto con una de nuestras creencias.

entre sí, al tiempo que ignoran al resto. Y, en la mayoría de los casos, las discusiones degeneran entre conflictos espirales entre extremistas de uno u otro punto de vista, lo que favorece la polarización o radicalización. La *homofilia*[9] *"o atracción entre iguales"* es la fuerza centrífuga que crea y cohesiona grupos ideológicos en las redes sociales.

Más a menudo, éstas trataran de operar sobre los sentimientos de los públicos objetivos (*target*) señalados, con finalidades políticas, es decir, trataran de generar con la información distorsionada estados de ánimo que difundidos en cascadas pueden generar rápidos estados de opinión de base emocionalmente inducida entre los grupos ya polarizados. Debemos recordar con (López Rosetti: 2018)[10] que los seres humanos no somos seres prioritaria o fundamentalmente racionales -si así fuera seriamos máquinas de procesamiento de información análogas a los ordenadores- sino, seres emocionales que razonan, que obviamente no es equivalente. Durante los cientos de miles de años de evolución de nuestra especie no existió la razón ni el pensamiento racional, solamente existían las emociones primero y los sentimientos después y hace muy poco tiempo, evolutivamente hablando, llego a través de la evolución el pensamiento racional con el lenguaje. Así lo señalan los brillantes estudios de (Haidt: 2019[11]). En efecto, el procesamiento de la información emocional no es ni Jeffersoniano, ni Platónico, es Humeano. Los seres humanos hacen juicios morales de manera rápida y emocional y el razonamiento moral es, principalmente, una búsqueda *post hoc* de razones para justificar los juicios que ya se han hecho. Lo anterior es consistente con los hallazgos de (Khaneman: 2012) en relación con el Sistema 1 y el Sistema 2 de procesamiento de información cerebral[12].

Lo anterior nos determina y, por ello, las emociones y la apelación a las mismas despliegan unos efectos, fundamentalmente en lo que respecta a las respuestas de temor, que la razón no es capaz de procesar en la misma forma que las emociones que son instantáneas e intensas, singularmente cuando éstas son audiovisuales tal y como ha estudiado (Amit, Hoeflin y colaboradores: 2017)[13] Estos aspectos son centrales para poder desarrollar en el futuro una regulación jurídica acorde con el impacto que sobre

---

[9] González de la Garza, Luis Miguel, "*Redes Sociales, Instrumentos de participación democrática. Análisis de tecnologías implicadas y nuevas tendencias*", Dykinson, Madrid, 2015.

[10] López Rosetti, Daniel, "*Emoción y sentimientos*", Ariel, Barcelona, 2018.

[11] Haidt, Jonathan, "*La mente de los justos. Por qué la política y la religión dividen a la gente sensata*", Deusto, Barcelona, 2019, págs.55-82.

[12] Khaneman, Daniel, "*Pensar rápido, pensar despacio*", Debate, Barcelona, 2012, pág. 33-84.

[13] Elinor Amit, Caitlyn Hoeflin, Nada Hamzah, y Evelina Fedorenko, "An asymmetrical relationship between verbal and visual thinking: converging evidence from behavior and fMRI", *Neuroimage*. 2017, mayo, nº 15; 152: págs. 619-627

nuestro sistema de procesamiento de la información tienen las "imágenes" en que consisten las Deepfakes, más adelante volveremos sobre esto.

Aunque Internet, como argumentan (Pluviano y Della Sala: 2019)[14]garantiza el acceso a una cantidad impensable de información hace veinte años, también ha difundido la idea de que basta utilizar un buscador y confiar en la veracidad de los resultados para hacerse un experto en un tema. Este efecto cognitivo se conoce como el efecto *Dunning-Kruger*. En 1999 los psicólogos sociales Justin Kruger y David Dunning describieron la incapacidad metacognitiva de las personas inexpertas de reconocer las propias limitaciones y su tendencia a sobrevalorarse "todos somos idiotas seguros de nosotros mismos" sostiene Dunning. Con frecuencia, el ignorante no sabe que lo es. En cambio, tiene una confianza exagerada e ilusoria en sí mismo, de manera que cree estar en posesión de la verdad. Se trata de una autosugestión. Para entender un argumento es indispensable analizar con sentido crítico aquello de lo que se lee. Por esa razón es necesario recurrir a expertos que ayuden a diferenciar hechos e informaciones falsas que difunden personas incompetentes más o menos conscientes de ello.

Lo anterior a nivel individual debe ponerse en conexión con algunos hallazgos epistemológicos recientes a nivel de grupo de ideología compartida (Marks, Copland y otros: 2019)[15] entre los que se encuentran aquellos que tienen que ver con el efecto cognitivo "*Halo*" por el cual se asigna a una persona o grupo competente en un área concreta y definida la misma competencia en áreas genéricas y generales formando un "halo de competencia" completamente injustificado y que induce a errores sistemícos entre quienes idealizan las comunicaciones de esas personas o grupos tomando la parte por el todo. La creencia generalizada entre los grupos en que "todo o casi todo" lo que proviene de un grupo es siempre cierto es uno de los efectos que conducen a públicos que se polarizan, un error asociado a la existencia de escasos datos y elevada incertidumbre en la emisión de un juicio sobre circunstancias o personas, lo anterior conduce a que las personas eligen escuchar a quienes tienen una opinión política similar *sobre temas que no tienen nada que ver con la política*, en lugar de a aquellas personas que saben que tienen información cierta sobre hechos relevantes, pero que tienen puntos de vista políticos distintos. Por ello dentro de los grupos una información falsa pero proporcionada por alguien en quien se confía puede ser difundida al encontrarse creíble por esa sola circunstancia.

[14] Pluviano, Sara, Sergio Della Sala, "El autoengaño de los antivacunas", en *Paradojas de la razón, Mente & Cerebro*, Investigación y Ciencia, marzo/abril 2019, nº 95, págs. 19.

[15] Marks, Joseph, Eloise Copland, Eleanor Loh, Cass R. Sunstein y Tali Sharot, "Epistemic spillovers: Learning others' political views reduces the ability to assess and use their expertise in nonpolitical domains", *Cognition* 188, 2019, págs. 74-84.

La democracia de opinión es lábil es un delicado mecanismo de equilibrios que opera o debe operar sobre la veracidad para que las elecciones de los ciudadanos se edifiquen sobre unos estándares mínimos de confianza legítima en el sistema de información. Es cierto que la naturaleza de Internet hace posible que las fuentes de devaluación de la información obedezcan en gran medida a actividades externas a la jurisdicción de cada Estado, por ello, una regulación global parece el mecanismo idóneo para limitar en la medida de lo posible esta grave amenaza disruptiva. Lo anterior ha sido adecuadamente recogido en el estudio "Desinformación en el ciberespacio" preparado por el Centro Criptológico Nacional[16] en el que se detalla cómo la desinformación es hoy un arma de primera magnitud en el arsenal de herramientas ofensivas de los Estados tecnológicamente avanzados, otras, por ejemplo, pueden ser los virus informáticos como lo fue Stuxnet, una de las primeras "ojivas electrónicas" destinadas a causar daños físicos en los sistemas informáticos. La única forma efectiva de respuesta eficaz a estos desafíos es crear una Internet intervenida por el Estado frente a tales ataques y que pueda desconectarse puntualmente de otros segmentos internacionales de red, que es exactamente lo que se ha hecho Rusia mediante el proyecto de reforma "sobre la introducción de enmiendas a algunos actos legislativos de la Federación Rusa" Artículo 1 modificación de la Ley Federal de 7 de julio de 2003 No. 126-FZ "0 Comunicaciones", actualmente Ley Federal de fecha 01.05.2019 No. 90-ФЗ "Sobre las Enmiendas a la Ley Federal Sobre Comunicaciones y la Ley Federal Sobre Información, Tecnologías de la Información y Protección de la Información "[17].

---

[16] Desinformación en el ciberespacio, "CCN-CERT BP/13", febrero de 2019, puede verse en: https://www.ccn-cert.cni.es/informes/informes-ccn-cert-buenas-practicas-bp/3549-ccn-cert-bp-13-desinformacion-en-el-ciberespacio.html

[17] http://publication.pravo.gov.ru/Document/View/0001201905010025

El proyecto de ley prevé las siguientes disposiciones. Se determinan las reglas necesarias para el enrutamiento del tráfico, se organiza el control sobre su cumplimiento. Crea una estrategia para minimizar la transferencia al extranjero de datos intercambiados entre usuarios rusos. Se definen líneas de comunicación transfronterizas y puntos de intercambio de tráfico. Sus propietarios, los operadores de telecomunicaciones, están obligados a garantizar la posibilidad de un control de tráfico centralizado en caso de una amenaza. Es posible instalar equipos técnicos en redes de comunicación que determinan la fuente del tráfico transmitido. Las instalaciones técnicas deberán poder restringir el acceso a recursos con información prohibida no solo por direcciones de red, sino también prohibiendo el paso de tráfico. Operabilidad de los recursos de Internet rusos en caso de imposibilidad de conectar operadores de telecomunicaciones rusos a servidores de Internet raíz extranjeros. Introduce la necesidad de ejercicios regulares de las autoridades gubernamentales, operadores de telecomunicaciones y propietarios de redes tecnológicas para identificar amenazas y desarrollar medidas para restaurar el segmento ruso de Internet. El Gobierno de la Federación de Rusia determina la respuesta centralizada a las amenazas a la operatividad de Internet y las redes de comunicaciones públicas por parte del Centro de Monitoreo y Gestión. Las medidas de respuesta se determinan, entre otras cosas, en el curso del monitoreo del funcionamiento de los elementos técnicos de la red pública de telecomunicaciones. Se está creando una infraestructura para habilitar las posibilidades que se definen en el paquete normativo. Esto es una circunstancia que sería algo semejante a revertir una forma de censura previa preventiva y a criterio del Estado, la motivación de la norma rusa se preparó, tal y como señala el propio texto de la norma, teniendo

2. ¿EN QUÉ CONSISTEN LAS DEEPFAKES?

En el año 2017, investigadores de la Universidad de Washington en los Estados Unidos demostraron cómo se podía manipular la tecnología de edición de generación de imágenes dinámicas, utilizando un algoritmo de aprendizaje profundo para imitar las expresiones faciales y la voz del presidente Obama, creando videos del expresidente que parecen hacer discursos con palabras de entrevistas anteriores[18]. Sin embargo, la generación de imágenes falsas tuvo parte de su origen, que no vamos a considerar aquí, en el uso de imágenes reales de actrices trasplantadas a cuerpos de otras actrices del mercado de la pornografía en redes como Reddit.

La tecnología de inteligencia artificial que hace posible la creación de las "Deepfake" posibilita crear videos sofisticados tan realistas que son casi imposibles de distinguir de la realidad. Las mentiras profundas o deepfakes son preocupantes precisamente porque permiten la manipulación de la imagen de cualquier persona u objeto y ponen en tela de juicio nuestra capacidad de confiar en lo que vemos. Un uso obvio de las deepfakes sería implicar falsamente a personas en escándalos de las más variadas naturalezas: desde los de carácter político; financieros, sexuales, etc. Otras opciones pueden basarse en reconstruir escenas históricas inexistentes como, por ejemplo, hacer decir a personajes históricos relatos que nunca se existieron, podríamos imaginar declaraciones de Hitler negando el holocausto y que éste fuese una estrategia aliada para denigrar al pueblo alemán, los ejemplos no tienen límite, pero sus efectos, como se puede apreciar serían profundamente disruptivos. Incluso si se demuestra que las imágenes incriminatorias o supuestamente históricas son falsas, el daño a la reputación de la víctima puede ser imposible de reparar o a la credibilidad de la historia difícil de mantener. Por ejemplo, los políticos podrían recrear viejas imágenes de sí mismos para que pareciera que siempre habían apoyado una narrativa política de conveniencia que recientemente se habría hecho popular, actualizando falsamente o recreando sus posiciones políticas en tiempo real, igualmente, sería posible generar de la nada imágenes ficticias de un político en actividades que nunca existieron.

Incluso podrían diseñarse figuras públicas o privadas que son completamente imaginarias, originales, pero no auténticas, es decir sintéticas o imágenes de síntesis. Mientras tanto, las imágenes de video podrían volverse inútiles como evidencia en los

---

en cuenta el carácter agresivo de la Estrategia Nacional de Seguridad Cibernética de los Estados Unidos adoptada en septiembre de 2018.

[18] https://www.washington.edu/news/2017/07/11/lip-syncing-obama-new-tools-turn-audio-clips-into-realistic-video/

tribunales. Las noticias de difusión audiovisual podrían reducirse a las personas que debaten si los videos son auténticos o no, utilizando una inteligencia artificial cada vez más compleja para tratar de detectar este tipo de mentiras profundas que no siempre serán capaces de detectar las manipulaciones mejor elaboradas (Bunk y colaboradores: 2017)[19] o (Li y Siwei: 2018)[20] entre otros. Existen, diversos tipos de deepfakes, desde aquellas que constituyen swaps de rostros (intercambio), las deepfakes de audio que imitan una forma de voz, las recreaciones faciales dinámicas completas, o aquellas que sincronizan los labios de la voz falsa y la insertan en un rostro público que reproducirá miméticamente la expresión facial fundida con la oral falsificada lo que desarrollará modelos miméticos de carácter sintético que recreara elementos de la comunicación no verbal capaces de generar la sensación en el auditorio que corresponde a la persona real de la que se falsifica esa información gestual característica, entre las más comunes.

Lo que está en juego con la aparición de estas falsificaciones de video profundas es la estructura social subyacente en la que la mayor parte de la sociedad, en un momento dado, está de acuerdo en que existe alguna forma de verdad mutuamente aceptada y ampliamente difundida y las realidades sociales que se basan en esta confianza. Esa es en esencia la base en la que se fundamenta el apartado d) del artículo 20 de la Constitución cuando establece que se reconocen y protegen los derechos a: enviar o recibir libremente información veraz. Esa veracidad entendida aquí como una forma de garantizar la verdad de la información objetiva -o como correspondencia con los hechos- es un elemento central de una opinión pública que recibe información tendencialmente verdadera. No se trata obviamente del fin de la verdad, sino del fin de la creencia en la verdad: una sociedad basada en la desconfianza es a lo que puede conducir un desarrollo masivo de las *deepfakes*. Tras la desinformación masiva, incluso las figuras públicas honestas podrían ser fácilmente ignoradas o desacreditadas. Las organizaciones tradicionales que han apoyado y permitido el consenso social y político, el gobierno y la prensa, ya no serán suficientemente aptas para el propósito que habrían venido desarrollando en el entorno no digital. Como señala (Frankfurt: 2007)[21] las ideas de verdad y facticidad son indispensables para dotar de plena sustantividad el ejercicio de la racionalidad.

---

[19] Jason Bunk, Jawadul H. Bappy, Tajuddin Manhar Mohammed, Lakshmanan Nataraj, Arjuna Flenner, B.S. Manjunath, Shivkumar Chandrasekaran, Amit K. Roy-Chowdhury, y Lawrence Peterson, "*Detection and Localization of Image Forgeries using Resampling Features and Deep Learning*", 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). arXiv:1707.00433v1 [cs.CV] 3 Jul 2017.

[20] Li Yuezun y Siwei Lyu, "*Exposing DeepFake Videos by Detecting Face Warping Artifacts***"** eprint arXiv:1811.00656, noviembre de 2018.

[21] Frankfurt, Harry G, "*Sobre la charlatanería y Sobre la verdad*", Paidós Contextos, Barcelona, 2007, pág. 107.

Algunas personas cuestionan los hechos en torno a eventos que sin duda sucedieron, tales como el Holocausto, el aterrizaje del hombre en la luna o los atentados del 11-S o que la tierra sea redonda, a pesar de las pruebas de toda clase existentes. Si las mentiras profundas logran que las personas crean que no pueden confiar en las imágenes, los problemas de la desinformación y las teorías de conspiración podrían empeorar significativamente como señalan (Jolley y Douglas: 2017)[22]. El pensamiento conspirativo se caracteriza, como señalan (Bauer, Bradley y Bangerter: 2013)[23] por la incapacidad de asignar a los hechos adversos un *determinante causal*, lo que implica un modo "*casi religioso*" de pensar en los procesos. Si bien, es cierto que la tecnología deepfake no es por el momento lo suficientemente sofisticada como para simular eventos o conflictos históricos a gran escala. Preocupa que la duda planteada por una o varias deepfakes convincentes y bien difundidas a escala internacional pueda alterar nuestra confianza en el audio y el video de forma permanente.

También son cada vez más fáciles y más baratos de crear los programas para la elaboración y desarrollo de este tipo videos, lo que significa que pronto será posible que cualquier persona con un ordenador personal y la capacidad apropiada de procesamiento -lo que no excluye procesamiento en red por grupos para disponer de mayor capacidad de cálculo- y el software adecuado, tengan los medios necesarios para crearlas y difundirlas de forma acelerada y eficiente en cualquier parte del mundo.

Es cierto que se existe una acción cada vez más acentuada para intentar detectar este tipo de mentiras profundas, así se puede observar el esfuerzo del proyecto liderado por DARPA[24] Media Forensics (MediFor). El programa Media Forensics está desarrollando herramientas capaces de identificar cuándo los videos y las fotos han sido alterados significativamente de su estado original para cambiar su contenido, el programa viene desarrollándose desde el año 2015 que se evidenció como problema de seguridad nacional en los Estados Unidos.

Señalaba Richard Feynman que *la ciencia es una larga historia de aprender la manera de no engañarnos*, quizá por ello necesitaremos desarrollar nuevas formas de consenso, nuevas formas de ponerse de acuerdo sobre situaciones sociales basadas en formas alternativas de confianza. Un enfoque prometedor, pero no exento de

[22] Jolley, Daniel y Karen M. Douglass, "The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprint", *British Journal of Psychology*, febrero de 2014.

[23] Frank, Bradley, Adrian Bangerter y Martin W. Bauer, "Conspiracy theories as quasi-religious mentality: an integrated account from cognitive science, social representations theory, and frame theory", *Frontiers in Psychology*, Julio, 2013.

[24] La Agencia de Proyectos de Investigación Avanzada de Defensa es la organización central de investigación y desarrollo del Departamento de Defensa de los Estados Unidos. http://www.darpa.mil/

limitaciones, advirtamos, podría ser descentralizar la confianza de modo que ya no necesitemos algunas instituciones clásicas para garantizar si la información es genuina, papel que ha venido siendo tradicionalmente desempeñado por la prensa o la televisión y la radio en sus diversas dimensiones, tanto en formatos materiales como inmateriales. Y, en cambio, se puede proponer confiar en redes de personas u organizaciones con buena reputación, pensemos como modelo en la Wikipedia, una enciclopedia virtual tan rigurosa (Giles: 2005)[25] como la enciclopedia británica. Una forma de hacer esto podría ser mediante el uso del blockchain, la tecnología que impulsa Bitcoin y otras criptomonedas. Blockchain funciona creando un libro de contabilidad público almacenado en varias computadoras de todo el mundo a la vez y a prueba de manipulaciones mediante la criptografía. Sus algoritmos permiten a las computadoras acordar la validez de cualquier cambio en el libro de contabilidad, lo que hace que sea mucho más difícil registrar información falsa. De esta manera, la confianza se distribuye entre todas las computadoras que pueden escrutarse mutuamente, aumentando la responsabilidad y haciendo posible, en hipótesis, construir mecanismos de verificación y contraste de fuentes y hechos de fiabilidad contrastada. De forma paralela es preciso construir sociedades más "*resilientes*" contra las deepfakes y la desinformación, capaces de poner en duda ese tipo de información cuando sea identificada. Antes de proseguir, es preciso identificar el concepto de "Resiliencia" que constituye el Pilar II de la estrategia de Cyberseguridad de la Casa Blanca[26]. El concepto de Resiliencia tiene su origen en el campo de la física y de la ingeniería. En tal sentido, se entiende por Resiliencia la magnitud que cuantifica la cantidad de energía que un concreto material puede absorber o almacenar al deformarse elásticamente pudiendo romperse o recuperarse de dicha deformación producida por efecto de un impacto por unidad de superficie de rotura. La Resiliencia se distingue de la tenacidad en que ésta cuantifica o mide la cantidad de energía absorbida por unidad de superficie de rotura bajo la acción de un esfuerzo progresivo y no por impacto, como en el caso de la Resiliencia[27]. En ingeniería de redes el término representa una forma de considerar la seguridad basada esencialmente en crear modelos de seguridad robustos y flexibles, de naturaleza proactiva y resistentes a diversos tipos de amenazas y ataques, capaces de generar daños de diversa magnitud en los elementos hardware y software que forman parte de la red, con capacidad de recuperación.

---

[25] Giles, Jim, "Internet encyclopaedias go head to head. Jimmy Wales, Wikipedia comes close to Britannica in terms of the accuracy of its science entries, a Nature investigation finds", *Nature*, vol 438, 15 de diciembre de 2005, pigs. 900-901.

[26] https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf (página 14, última visualización 5 de marzo, 2019)

[27] La Resiliencia se expresa como unidad de medida en Julios por metro cuadrado (J/m^2)

Convendría señalar brevemente algunos aspectos fundamentales sobre el tipo de red que es Internet, es decir, como red con topología libre de escala. Siguiendo los trabajos a los que remitimos, de (Albert y Barabási: 2000)[28], (Strogatz y Watts: 2008)[29] y, más recientemente y con una articulación más asequible al lector no experto en redes, de (Ball: 2009)[30], trabajo altamente ilustrativo para comprender la estructura de las redes y, en particular, los aspectos esenciales sobre la seguridad que vertebra la topología de este tipo de redes. Internet (como red física de enlaces) y la Web (como red de conexiones lógicas) son redes de topología libres de escalas con una distribución de conectividad entre dos nodos que obedece a una ley de potencias[31]. En esto reside su fuerza ya que las redes libres de escala son, en general, resistentes a fallos aleatorios (accidentales)[32], siendo por ello redes altamente flexibles a ese tipo de incidentes. A diferencia de una creencia común, la red Internet posee una estructura que, podríamos denominar, "grumosa" o de "grumos". Existen enlaces muy conectados y otros, la inmensa mayoría, débilmente conectados. Eso no significa que no exista redundancia entre los enlaces. Existe, desde luego, y ésta es precisamente la que permite la reorganización de la estructura tras las situaciones de desorden, debidas a colapsos temporales y puntuales. Pero la redundancia se expresa, con mucha mayor intensidad, en los puntos o nodos más densamente conectados de las redes. Aquí reside exactamente la debilidad de las redes sin escala frente a acciones "premeditadas" de ciberterrorismo y desarrollo de cascadas de información viral como las fake news. Cuando los nodos más conectados de Internet o de la Web son destruidos, las redes de topología sin escala se vienen abajo rápidamente. Destruir uno de cada 20 nodos de alta densidad de conexiones duplica la longitud media -en distancia física y lógica- de las

[28] Albert, R y Barabási, A.L, "Topology of evolving networks: Local events and universality", *Physical Review Letters*, nº 85, 2000 págs. 5234-5237.

[29] Watts, D.J. y S.H. Strotagtz, "Collective dynamics of *small-world* networks", *Nature*, nº 393, 1998, págs. 440-442.

[30] Ball, Philip, "*Masa crítica, cambio, caos y complejidad*", Turner/FCE, México, 2009, págs. 437-471.

[31] La peculiaridad que caracteriza a las redes libres de escala es que los nodos que conecta la red carecen de una distribución homogénea -*sin escala*-, sin embargo, en estas redes, estadísticamente, se comprueba en los estudios de agregación una *ley de potencias* que las caracteriza. Existen nodos ultradensos o muy relacionados y otros que no lo están. Tal hecho fue observado por Wilfredo Pareto, en el marco de las redes sociales, en 1890 cuando el autor advirtió que el 20% de la población era la dueña del 80% de la riqueza del país, originando la conocida regla 80-20. Posteriormente, en 1940, Geork K. Zipf observó un fenómeno análogo en el uso estadístico de las palabras de un lenguaje, es decir, al escribir se emplea mucho un pequeño número de palabras mientras que la mayoría de las palabras no son usadas. Este hallazgo se tradujo en la denominada Ley de *Zipf*.

[32] Es decir, fallos no planificados por la inteligencia humana, tales como congestiones puntuales de redes por excesivo tráfico, caídas de segmentos de red por averías puntuales que obligan a las redes autónomamente a reconfigurar su carga de trabajo y distribución, etc.

conexiones de inmediato, con lo que para llegar de un sitio a otro de la red son necesarios grandes rodeos. Como señala (Watts, 2006)[33], las jerarquías, como podemos figurarnos, responden muy mal en condiciones de avería, fallo o colapso. Por la misma razón que son vulnerables a los colapsos y a los fallos relacionados con la congestión (debido a que son demasiado centralizadas), si cualquiera de los nodos superiores de una jerarquía falla, grandes trozos o segmentos de la red quedarán aislados unos de otros. Así, por ejemplo, si un ciberterrorista o conjunto de éstos, en acciones de ataque distribuidas pero organizadas quieren destruir temporalmente la conectividad de Internet, les basta con identificar un número relativamente pequeño de los nodos más concentrados y sabotearlos. Como argumenta (Ball, 2009)[34], cuando la inteligencia criminal guía los atentados, una red sin escala como Internet puede sufrir daños desproporcionados. Por ello, es alrededor de los nodos más conectados donde habría que erigir las murallas defensivas más seguras. El talón de Aquiles de las redes sin escala es la enorme influencia de algunos nodos muy conectados: los núcleos que mantienen la coherencia del conjunto. Si alguien interrumpe los contactos o conexiones de estos importantes núcleos (los grumos) la Web entera se desploma rápidamente.

Vemos pues que la Resiliencia debe operar sobre un entorno -brevemente descrito- en extremo complejo y que la misma, en nuestra opinión, debe articularse en los núcleos centrales de interconexión. Existen otros conceptos de Resiliencia como los empleados, por ejemplo, en la Psicología y que tienen también una función explicativa de interés. Resumamos tan sólo algunos de tales conceptos. Así, la Resiliencia sería un proceso dinámico que tendría por resultado la adaptación positiva en contextos de gran adversidad, o la habilidad para resurgir de la adversidad, adaptarse, recuperarse y acceder a una vida significativa y productiva. Como podemos observar, se trata de una traslación del significado en Física y en Ingeniería, transformando la capacidad de deformación elástica de los materiales y sustituyéndola por la capacidad de adaptación de los seres humanos frente a cargas estresantes y desestabilizadoras de fuerte impacto emocional (enfermedad, muerte, desempleo, violencia, divorcio, o un entorno de comunicaciones con abundantes deepfakes).

Estamos empezando a disponer cada vez de mejores y más precisos modelos para el estudio de cómo se difunden sobre las redes IP las mentiras, tanto las fake news como las deepfakes, así los trabajos de (Kopp, Korb y Mills: 2018)[35] son de gran valor,

---

[33] Watts, Duncan J, "*Seis grados de separación. La ciencia de las redes en la era del acceso*", Paidós, Barcelona, 2006 pág. 287.

[34] Ball, Philip, "*Masa crítica, cambio, caos y complejidad*", 2009, Ibíd. pág. 457.

[35] Kopp, Carlo, Kevin B. Korb y Bruce I. Mills, "Information-theoretic models of deception: Modelling cooperation and diffusion in populations to "fake news", *Plos One*, 28 de Noviembre de

singularmente, el modelo de engaño de Borden-Kopp en el que se definen cuatro modelos de teoría de la información, *degradación*, *corrupción*, *negación* y *subversión*, cada uno de los cuales es una forma específica de alterar la percepción de la víctima. Como se observa empíricamente en las simulaciones de agentes realizadas por los investigadores, en los sistemas sociales sometidos a ataques de "noticias falsas" sean estas fake news o deepfakes, incluso en los casos en los que se trata de una población muy pequeña de engañadores que aparecen transitoriamente, estos pueden alterar el equilibrio de la población en favor de las estrategias de explotación, a expensas de las estrategias de cooperación. Los resultados también muestran que la capacidad de una población de engañadores para establecerse o permanecer presente en una población es altamente sensible al costo del engaño, ya que este costo reduce la aptitud de los engañadores cuando compiten contra agentes que no engañan, es decir, en los modelos el castigo disminuye sensiblemente la existencia del engaño.

3. LAS REDES NEURONALES GENERATIVAS ANTAGÓNICAS

La inteligencia artificial[36] y en particular las redes neuronales antagónicas (GAN) como señala (Condliffe: 2018) están siendo progresivamente capaces de identificar cosas u objetos con gran precisión, es factible mostrarles diez millones de fotografías y éstas, las GAN, podrán identificar con asombrosa precisión en cuáles de ellas aparece, por ejemplo, una persona montando en bicicleta circulando por una calle. El problema es que para crear algo completamente nuevo hace falta imaginación, circunstancia que hasta ahora no era posible en virtud de los modelos y procedimientos disponibles en inteligencia artificial. La primera solución la propuso el, entonces estudiante de doctorado, en la Universidad de Montreal (Canadá) Ian Goodfellow en el año 2014[37]. No debemos olvidar sin embargo que uno de los trabajos seminales en redes neuronales fue el de (Hinton, Rumelhart y Williams: 1985)[38] ese trabajo desarrolló una técnica llamada

2018, puede verse en: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0207383 (consultado el 02/01/2019)

[36] Condliffe, Jamie, “Redes Generativas Antagónicas”, MIT Technology Review, 22 de febrero de 2018. https://www.technologyreview.es/s/10026/tr10-redes-generativas-antagonicas (última consulta, marzo 2019).

[37] Goodfellow, Ian J, Jean Pouget-Abadie, y otros, “Generative Adversarial Nets”, Departement d’informatique et de recherche operationnelle ´ Universite de Montreal, Montreal, QC H3C 3J7. Puede obtenerse en Cornell University: https://arxiv.org/abs/1406.2661

[38] Rumelhart, David E, Geofrey E. Hinton y Roanald J. Williams, “Learning internal representations by error propagation”, *ICS Report* 8506, Septiembre de 1985.

retropropagación que es en lo que se basa todo el aprendizaje profundo[39] incluidas las *nuevas redes recurrentes* (Bengio: 2020)[40].

El enfoque conocido como redes generativas antagónicas (GAN, por sus siglas en inglés) emplea dos redes neuronales -modelos matemáticos simplificados del cerebro como los descritos por (Hawkins y Blakeslee: 2005)[41]- que se enfrentan entre sí, es decir, la "*confrontación entre redes*" es la esencia del aprendizaje autónomo que se produce como producto de esa confrontación. Ambas redes están entrenadas con el mismo conjunto de datos. Una red conocida como la generativa tiene la tarea de crear variaciones en las imágenes que ya ha visto, tal vez una imagen de una bicicleta con una rueda de más. La segunda, conocida como *el discriminador* debe identificar si la imagen que está viendo pertenece al conjunto de entrenamiento original o, si, por el contrario, es una imagen falsa producida por la red generativa. A la red discriminadora básicamente se le formula la siguiente cuestión: ¿Es probable que una bicicleta con tres ruedas sea real? Con el tiempo, a la red generativa se le da tan bien producir imágenes que a su pareja discriminadora le resulta imposible detectar la falsificación. En resumen: la red generativa aprende a reconocer y posteriormente a crear imágenes de bicicletas de aspecto realista. Esta tecnología se ha convertido en uno de los avances más prometedores de la inteligencia artificial en la última década, capaz de ayudar a las máquinas a producir resultados que engañan incluso a los humanos expertos. Las GAN se usaron para crear sonidos e imágenes hiperrealistas. En un convincente ejemplo, los investigadores del fabricante de chips gráficos Nvidia entrenaron, como señala (Borel: 2018)[42] a una GAN con fotografías de personas famosas para que el sistema fuera capaz de crear cientos de rostros creíbles de personas que no existen. Otro grupo de investigación consiguió generar pinturas falsas parecidas a las obras de Van Gogh. Si se les fuerza aún más, las GAN pueden reinterpretar las imágenes de diferentes maneras: pueden hacer que una carretera soleada parezca nevada o convertir caballos en cebras.

Los resultados no siempre son perfectos, las GAN pueden crear bicicletas con dos tipos de manillar o tres sillines, por ejemplo, o caras con cejas en el lugar incorrecto del rostro humano. Pero debido a que las imágenes y los sonidos son, por lo general, extraordinariamente realistas, algunos expertos creen que hay una lógica detrás de cómo

---

[39] La expresión de "profundas" se debe a que tienen muchas capas de nodos de cálculos simples que trabajan en conjunto para buscar datos y entregar un resultado final en forma de predicción. Para observar las diversas topologías de redes neuronales, puede verse: http://www.asimovinstitute.org/neural-network-zoo/ (última visualización, marzo 2019).

[40] Bengio, Yoshua, "Aprendizaje profundo", *Temas*, Investigación y Ciencia, 1 trimestre 2020, Nº 99, págs.6-11.

[41] Hawkins, Jeff, Sandra Blakeslee, "*Sobre la Inteligencia*", Espasa, Madrid, 2005, págs. 37-55.

[42] Borel, Brooke, "Clics, mentiras y cintas de video", *Investigación y Ciencia*, diciembre 2018, págs.31-35.

las GAN comienzan a comprender la estructura subyacente del mundo que ven y oyen. Otros, en cambio a quienes nos adherimos piensan que tanto los algoritmos como las redes neuronales que emplean estas tecnologías, no son en absoluto conscientes de su propia existencia, carecen de inteligencia y de autopercepción o conocimiento de si mismas, características esenciales de la inteligencia humana, son emulaciones y simulaciones rudimentarias pero sencillamente no saben lo que hacen, circunstancia que si comprende quienes las diseñan. El potencial de las GAN es muy grande, porque pueden aprender a imitar cualquier distribución de datos. Es decir, se puede enseñar a las GAN a crear mundos inquietantemente similares a los nuestros en cualquier dominio: imágenes, música, habla, prosa, etc. Los tipos de redes GAN evolucionan de forma constante, por ejemplo, como señalan (Bansal y colaboradores: 2018)[43] las redes *recycle-GAN* son capaces de traducir el contenido de un dominio a otro, preservando el estilo nativo del primer dominio, es decir, si los contenidos del discurso de un participante se transfieren a otro, se transfieren con el estilo propio del primero al segundo en un proceso de aprendizaje profundo automatizado sin supervisión.

4. UNA APROXIMACIÓN A LA FORMA DE AFRONTAR JURÍDICAMENTE EL RETO DE LAS DEEPFAKES

Con estas fuentes de información cada vez más realistas, una de las preguntas que debemos formularnos es la de si son los instrumentos legales en vigor aptos para ofrecer respuestas jurídicas adecuadas a las deepfakes, por una parte y también parece pertinente preguntarnos si es prudente o aconsejable una respuesta preventiva que pueda limitar arbitrariamente la libertad de expresión.

Hay que empezar por señalar que se trata de un problema complejo con múltiples matices y dimensiones de análisis. En relación con la primera pregunta la respuesta inicial es que, sí disponemos de un instrumental jurídico que realizando adaptaciones legislativas apropiadas para ofrecer una respuesta proporcionada al nivel de amenaza de las deepfakes sea capaz de hacerlas frente. Tanto las disposiciones del Código Penal,[44] como las establecidas en la Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico, así como las previstas en la, Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, en Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, la Ley 17/2001, de 7 de

[43] Bansal, Aayush, Shugao Ma, Deva Ramanan y Yaser Sheikh, "Recycle-GAN: Unsupervised Video Retargeting", *ECCV,* 2018. http://www.cs.cmu.edu/~aayushb/Recycle-GAN/

[44] Davara Fernández de Marcos, Elena, Laura Davara Fernández de Marcos, "*Delitos Informáticos*", Aranzadi, Pamplona, 2017.

diciembre, de Marcas conjuntamente con las normas Civiles y Administrativas pertinentes, pueden ser suficientes inicialmente al menos para resolver algunos de los conflictos que este tipo de información puede generar. Seguramente donde sea preciso realizar modificaciones para lograr tiempos de respuesta adecuados a los daños que se puedan generar, sea en las normas procesales Penales, Civiles y Administrativas donde el tiempo de respuesta puede ser clave para la mitigación de daños.

Cuestión diversa sin duda puede ser la respuesta de los Cuerpos y Fuerzas de Seguridad del Estado cuando tales deepfakes por su intencionalidad, origen o características estén dirigidas a la desestabilización del país, en ese caso será de aplicación la Ley 36/2015, de 28 de septiembre, de Seguridad Nacional, así como el Real Decreto-ley 12/2018, de 7 de septiembre, de seguridad de las redes y sistemas de información, fundamentalmente.

Pensamos que desde la dimensión procesal considerada se puede hacer inicialmente una división de procesos ordinarios y especiales. Entre los ordinarios cabría considerar los de naturaleza civil y penal relacionados con la libertad de expresión, honor y propia imagen en contextos no especiales que veremos seguidamente. Recordemos previamente que lo audiovisual (Deepfake) es muy diferente en su interpretación mental a la interpretación de un texto escrito (Fake news). La forma cognitiva de tratamiento de la información por el cerebro humano es muy distinta. Las imágenes y el sonido, siguiendo a (Margolis:1987)[45] se procesarían emocionalmente según su descripción "*ver qué*" en lugar de como se procesaría lenguaje escrito que correspondería con "*Razonar por qué*", sistema descriptivo equivalente al Sistema 1 de Khaneman caracterizado aquí como un pensamiento *rápido, automático y sin esfuerzo.* Esa forma de procesamiento emocional de la información audiovisual implica una respuesta jurídica acelerada para limitar el alcance de los posibles daños de las Deepfakes. Es decir, es fácil que este tipo de informaciones audiovisuales puedan transformarse en virales por su componente emocional más que racional y el daño puede seguir, en estos supuestos, una progresión exponencial. El legislador ha de ser consciente de la naturaleza del tipo de información y la respuesta emocional para adaptar la respuesta jurídica. Nos parecen adecuadas aquí las órdenes de cesación dirigidas a los proveedores de servicios. Ahora bien, es razonable pensar que las medidas han de rodearse de garantías adecuadas para no favorecer una censura no justificada. Quizá para ello lo pertinente, aunque aún no existe, sería la creación de juzgados especializados en éste tipo supuestos ilícitos de base virtual en los que la decisión de retirada de la información se base en jurados también virtuales que puedan, en plazos inferiores a 6 horas, dar un veredicto inicial de retirada o

[45] Margolis, H, "*Petterns, Thinking and Cognition*", Chicago University Press, Chicago, USA, 1987.

mantenimiento de la información a propuesta de las partes que entiendan que esa información daña en alguna forma sus derechos con participación del Ministerio Fiscal y con presencia de quienes difunden la información sujeta a enjuiciamiento.

Entendemos por contextos especiales, inicialmente el electoral, en el que sería la Junta Electoral Central la que debería jugar el papel central en el enjuiciamiento de estas informaciones de naturaleza política, durante los plazos de campaña electoral, sin perjuicio de revisar plazos más amplios en los que la Junta tutele la pureza de la campaña y pueda velar por su ordenada realización en la dimensión en la que opera este tipo de información en las redes sociales.

Naturalmente aquí se plantea en alguna medida la tendencia hacia la distopía advertida por Georg Orwell en su célebre obra "1984" con la creación de un Ministerio de la verdad, junto con los de la Abundancia, La Paz y el Ministerio del Amor. Que se dedicaba sistemáticamente a modificar o destruir los documentos históricos de todo tipo: libros, revistas, películas…para lograr que las evidencias del pasado coincidan con la versión oficial de la historia mantenida por el Estado. Goethe señalaba que los hechos históricos son a veces homólogos, pero nunca análogos, Goethe no imaginaba la capacidad de falsificación de las evidencias del pasado por la tecnología presente. Quizá no se llegue a un Ministerio, pero no es impensable que organizaciones políticas podrían reescribir el pasado o una parte de él con las finalidades más espurias que podamos suponer. Eso sucede en la actualidad con ideologías nacionalistas y es un riesgo cierto que hay que advertir.

La mentira, como recuerda (Sunstein: 2019)[46] ha sido admitida por la doctrina del Tribunal Supremo de los Estados Unidos desde que en el año 2012 se dictara la sentencia *EE. UU vs Álvarez*[47] en amplios contextos y por algunas buenas razones que no vamos a considerar aquí y que se fundamentan en una interpretación de la primera enmienda de la Constitución norteamericana tal vez de forma excesivamente expansiva, recordemos que el tribunal estuvo muy dividido. Los argumentos de los jueces discrepantes, Samuel Alito, Antonin Scalia y Clarence Thomas hoy tendrían un significado de gran valor en lo que respecta a las Deepfakes y seguramente la mayoría también sería de diverso parecer en el contexto cultural actual. En los 7 años transcurridos desde la decisión, en términos de nuevas tecnologías es casi como si hablásemos de una generación, de 30 o 40 años.

[46] Sunstein, Cass R, "Falsehoods and the First Amendment", puede leerse en: SSRN: https://ssrn.com/abstract=3426765 or http://dx.doi.org/10.2139/ssrn.3426765 (consultado 1/09/2019)

[47] https://supreme.justia.com/cases/federal/us/567/709/

En relación con la segunda pregunta las deepfakes cuando por su origen e intencionalidad no tengan una finalidad claramente dirigida desde el exterior o desde el interior a la desestabilización del país, o a afectar en alguna medida al proceso electoral o tengan una intencionalidad delictiva, pensemos en una injuria o calumnia. En la inmensa mayoría de los casos se tratará de parodias que, inicialmente y bajo tal régimen habrán de ser consideradas y valoradas por la jurisdicción en los casos en los que por su forma o circunstancias de ejecución se residencien ante la misma en el marco de los conflictos de la propiedad intelectual. La parodia como señala (Sol Muntañola: 2005)[48]es un límite al derecho, limita los derechos del autor o, dicho de otra forma, permite a cualquiera utilizar libremente -aunque de acuerdo con una serie de condiciones legales- una obra protegida por el derecho de autor. En este sentido, la parodia invade el derecho dominical del autor, apartándole de su señorío, para destruir y reutilizar su creación. Ciertamente este límite es quizá uno de los más importantes en relación con las deepfakes. Valorar cuando determinados contenidos se encuentran bajo la órbita expansiva, en una sociedad democrática, de la libertad de expresión será una tarea delicada. La Directiva de Derechos de autor y mercado único digital en sus últimas fases de elaboración reconoce este derecho en el apartado 5 de su artículo 13, pero surgen dudas en relación con la elaboración de una deepfake que sea el producto de una imagen de síntesis pero que no es una exacta versión de una imagen original de la persona sobre la cual se genera el video. Los programas de edición de video como *Adobe After Efects*, pueden generar una deepfakes, pero no son el producto de elaboración de las GAN que anteriormente consideramos, existen diversas novedades en la generación de la imagen de composición y de una imagen real pero distorsionada por edición, que será preciso discriminar para la aplicación de las excepciones.

En los Estados Unidos los límites actuales frente a las deepfakes son básicamente el derecho de autor, el derecho de publicidad, la sección 43 (a)[49] de la Ley Lanham, las normas penales por daños, por difamación y las de extorsión en los casos y supuestos en los que una deepfake pudiese utilizarse con tal finalidad. Son igualmente de

---

[48] Sol Muntañola, Mario, "*El régimen jurídico de la parodia*", Marcial Pons, Madrid, 2005, pág. 102.

[49] La Sección 43 (a) de la Ley Lanham proporciona dos teorías generales de responsabilidad:

(1) representaciones falsas con respecto al origen, respaldo o asociación de bienes o servicios mediante el uso indebido de la marca distintiva, el nombre, la imagen comercial u otro dispositivo (" endoso falso " o " asociación falsa ") de otra persona, y

(2) Representaciones falsas en publicidad sobre la calidad de los servicios o productos (" publicidad falsa ").

"Para prevalecer bajo la sección 43 (a) de la Ley Lanham, un demandante debe demostrar que tiene una marca comercial válida y protegida y que el uso por parte del demandado de una imitación semejante de la marca comercial puede causar confusión entre los consumidores".

aplicación la *Intentional infliction of emotional distress* (IIED) en los casos de conductas escandalosas las cuales podrían afectar a políticos o personajes públicos siendo éste *Tort of outrage* de aplicación en diversos supuestos. A nuestro juicio y lo que sólo puede ser un esbozo en este trabajo, las normas actualmente disponibles son suficientes para ofrecer una respuesta en Derecho a los daños que pueda originar este tipo de informaciones de video. Por otro lado, y dado que el fenómeno se está definiendo en estos momentos tampoco se puede afirmar que sean insuficientes, es preciso seguir evaluando las formas de evolución del fenómeno sin que se limite la libertad de expresión por una tecnología que, como tal, ni es ilícita ni muestra ningún tipo de maldad intrínseca.

La tecnología no es mala, lo puede ser, como de costumbre a lo largo de la historia de los avances técnicos, los usos que de ella se hagan. Aun así, los legisladores se sienten bajo la presión de adoptar medidas contingentes en situaciones de incertidumbre social pero todas ellas deben analizarse serena y cuidadosamente porque lo que pueden cercenar es la libertad de expresión. Muestra de lo anterior y siendo formas de precipitación legislativa puede considerarse el fallido proyecto de Ley presentado en el Senado de los EE. UU., S.3805, de prohibición de ciertos registros audiovisuales fraudulentos y para otros propósitos, presentado por el Senador por Nebraska Ben Sasse.

5. LA PROPAGANDA COGNITIVA ELECTORAL Y LA MICROSEGMENTACIÓN

La propaganda computacional es ya una realidad como precisan entre otros (Wooley y Howard: 2017) si bien lo novedoso del tipo de propaganda que estamos examinando no es que sea una propaganda pasiva, sino que es una propaganda que podríamos denominar activa e inteligente debido a que se aprovecha de los sesgos caracterológicos de los votantes para diseñar una campaña de muy alta granularidad y precisión a la medida del elector y de sus preferencias emocionales y políticas. Si, por ejemplo, es un elector que se ha abstenido en otras elecciones pueden ofrecérsele argumentos basado en sus preferencias emocionales conocidas para que vote. Podemos pensar en votantes que expresen caracteres que puedan ser explotables por agentes de propaganda automatizada, votantes que no tienen una clara preferencia y a los que este tipo de propaganda puede “seguir” de forma que mediante el “microtargeting o microsegmentación” éste busque al elector para ofrecerle propaganda activa de su agrado, capaz de aprender de la interacción con el votante en base a su personalidad y readaptarse y refinarse en función de las respuestas del votante a un diálogo virtual de acompañamiento propagandistico que con anterioridad al advenimiento de estas

tecnologías era inexistente. Es decir, Deepfakes dirigidas a explotar emotivamente los sesgos de su carácter.

Se denomina microtargeting porque tiene por objetivo agrupar a los electores en muy pequeños segmentos o *clusters*[50]sincronizados con los 20 modelos de tipos de personalidades o perfiles psicométricos ya elaborados a los que se dirige este tipo de propaganda electoral. Para que la información personalizada alcance y siga a su objetivo electoral. Es usual observar en cualquier navegación por Internet que tras visitar un comercio virtual posteriormente aparece en nuestros ordenadores o teléfonos móviles información del producto o servicio que hemos visitado anteriormente, en horas, días o semanas anteriores, la publicidad sigue al navegador en determinadas páginas Web merced al uso de cookies[51] previamente aceptadas e instaladas en los equipos de los usuarios en los que esta publicidad contextual "que nos busca y acompaña" aparece. Ese seguimiento sería el equivalente del microtargeting electoral en su dimensión comercial. Pero a diferencia de ese microtargetin comercial, el electoral interactúa y aprende del elector al que tratará de *persuadir* con argumentos *racioemocionales* intentando imitar los intereses personales, sociales y emocionales de éste y ofrecer al mismo, variantes de campaña propagandística adaptadas a su perfil psicológico. Los experimentos de manipulación y contagio masivo de emociones en las redes sociales como el que se produjo en Facebook el año 2012 y que afecto a 700.000 sujetos como estudiaron (Kramer, Guillory y Hancock: 2014)[52] demuestran convincentemente la gran efectividad de lo que se puede lograr en el ámbito de la transformación de motivaciones y preferencias mediante contagio emocional inducido.

En la campaña electoral de Donald Trump del año 2016, Cambridge Analytica, en la actualidad Emerdata tal y como señala (Cadwalladr: 2017)[53] estaba empleando entre cuarenta y cincuenta mil variantes de diferentes argumentos electorales informativos de los que se medía su respuesta en tiempo real de los destinatarios, readaptándose a sus respuestas de forma evolutiva. La granularidad de las acciones de estos mensajes está

---

[50] De tipo sociodemográfico como distritos electorales o circunscripciones específicas disputadas, en las que pocos votos pueden producir la asignación de un escaño y en los que una actividad de propaganda cognitiva puede justificar un esfuerzo suplementario de campaña electoral para conducir a votantes que dudan en si ejercerán o no su voto a ser motivados a decantarse hacia una tendencia electoral determinada.

[51] López Jiménez, David, "Las cookies como instrumento para la monitorización del usuario en la red: La publicidad personalizada", *Ciencias Económicas* 29, nº 2, 2011.

[52] Kramer, Adam D.I, Jamie E. Guillory y Jeffrey T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks", *PNAS*, Vol. 111, nº 24, 17 de junio 2014, págs. 8788-9790.

[53] Cadwalladr, Carole, "Google, democracy and the truth about internet search" Internet, The Observer, 4 de Diciembre de 2016. https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook (20 de Agosto de 2017)

estructurada por áreas geográficas de hasta una radio de 5 millas en las que se agrupan los perfiles psicográficos[54] que se evalúan por el algoritmo de Cambridge Analytica, cuyo origen se encuentra en la Universidad de Cambridge[55]. Además, las variantes de los mensajes propagandísticos empleados actualmente no pueden ser conocidos por otros electores ya que se basan, por ejemplo, en Facebook en las publicaciones invisibles[56]o *dark post* que inicialmente fueron y son un instrumento para la publicidad personalizada pero que también se puede emplear en las campañas electorales cognitivas personalizadas y que son de difícil fiscalización por una futura autoridad electoral.

(Wolley y Howard: 2017)[57] señalan y, nos adherimos a sus conclusiones, que la propaganda computacional es una de las herramientas más poderosas contra la democracia ya que hace posible una genuina y nueva forma de "ingeniería social" capaz de romper por completo los modelos de opinión pública y de su manipulación como han estudiado (Bond, Fariss y colaboradores: 2012)[58]. En efecto, los sistemas propaganda cognitiva electoral parece que funcionan en paralelo a poderosas y profundas distorsiones de la opinión pública que están siendo originadas por muy diversos grupos de interés de alcance nacional e internacional capaces de modificar, por ejemplo, mediante granjas de ordenadores la agenda de la opinión pública en temas de interés político mediante la manipulación de tendencias basadas en generación de hashtags hasta lograr posicionamientos como Trending Topics como señala (Nimmo: 2019)[59]. Si bien, esas tendencias son creadas de forma artificial e intencionada tanto por las señaladas granjas de ordenadores, como por *bots*[60] automatizados (Ferrara y otros: 2016)[61] u otros vectores tecnológicos de generación y difusión al servicio de sus

[54] La segmentación psicográfica es una herramienta que permite profundizar en los grupos de referencia para encontrar sus motivaciones de voto.

[55] El lector puede experimentar un análisis psicográfico básico de sus redes sociales con este algoritmo en: https://applymagicsauce.com/

[56] https://www.facebook.com/business/a/online-sales/unpublished-page-posts (20 de agosto de 2017)

[57] Woolley, Samuel C, y Philip N. Howard Op Cit, pág. 7.

[58] Bond, Robert M, Christopher J. Fariss, Jason J. Jones, AdamD.I.Kramer, Cameron Marlow, Jaime E. Settle& James H. Fowler, "A 61-million-person experiment in social influence and political mobilization", Nature, Vol 489, 13 Septimbre 2012, pag.295-298.

[59] Nimmo, Ben, "*Measuring Traffic Manipulation on Twitter*", Computational Propaganda Research Project, Oxford University, 2019.

[60] Para una taxonomía de los diversos tipos de Bots aptos para ingeniería social, puede verse: Ferrara, Emilio, Onur Varol, Clayton Davis, Filipo Menczer y Alessandro Flamini, "The Rise of Social Bots", *Communications of the ACM*, Julio 2016, Vol. 59, nº 7.

[61] Ferrara, Emilio, Onur Varol, Clayton Davis, Filipo Menczer y Alessandro Flamini, "The Rise of Social Bots", Communications of the ACM, Julio 2016, Vol. 59, nº 7.

creadores. El fenómeno ha sido estudiado por (Bradshaw y Howard: 2017)[62] en el contexto internacional hallándose un cuerpo de evidencias muy preocupante ya que la principal tarea de estas plataformas, que fue en su origen dar forma a la opinión pública a través del uso de "narrativas dinámicas" para combatir la propaganda diseminada en las redes por las organizaciones terroristas, ha cambiado en la actualidad alcanzado otras actividades completamente diversas como las de naturaleza política al demostrarse una efectividad o eficacia de estas técnicas en finalidades distintas de para las que fueron originariamente diseñadas, es decir, una herramienta diseñada contra el terrorismo se emplea ahora para las campañas electorales.

Lo señalado hace referencia a la elaboración de la información falsa, posteriormente esta información es difundida o vectorizada en las redes sociales por grupos o personas. Recientemente (Guess, Nagler y Tucker: 2019)[63] han estudiado qué grupos sociales -por edad- son los agentes de difusión más característicos en redes como Facebook, llegando a la conclusión de que un pequeño porcentaje de estadounidenses, menos del 8,5 por ciento compartió enlaces a los sitios de "noticias falsas" durante la campaña electoral de 2016, pero este comportamiento fue desproporcionadamente común entre las personas mayores de 65 años con independencia de la afiliación ideológica o política los jóvenes tuvieron un papel muy inferior.

6. LA REGULACIÓN EN ESPAÑA DE LAS CAMPAÑAS ELECTORALES COGNITIVAS VIRTUALES

Nuestro legislador ha regulado la inclusión de estas campañas electorales en un nuestro ordenamiento, cuando podía no haberlo hecho, dejando la regulación intacta lo que supondría la exclusión del ordenamiento jurídico de esos riesgos señalados. No solamente el legislador español ha regulado la materia, sino que la ha regulado de forma contradictoria y defectuosa. Por otra parte, la tramitación de la norma electoral se produce a través de la reforma de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, es decir, se lleva a cabo de forma indirecta lo que es una reforma muy profunda de las campañas electorales. Tal forma de proceder, si bien jurídicamente no es ilegítima, si representa en

[62] Bradshaw, Samantha y Philip N. Howard, "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation", Working paper nº 2017.12, Computational Propaganda Research Project, Oxford University, UK

[63] Guess, Andrew, Jonathan Nagler y Joshua Tucker, "Less than you think: Prevalence and predictor of fake news dissemination on Facebook", Sci. Adv. 2019; 5: eaau4586 9 January 2019. Puede consultarse en: https://advances.sciencemag.org/content/advances/5/1/eaau4586.full.pdf%20 (visto, 8 de mayo de 2019.

cambio, a nuestro parecer, la intención del legislador de ocultar a la ciudadanía una reforma tan importante por lo que se podría denominar, la puerta de atrás. Pensemos que la reforma se efectúa por medio de la disposición adicional tercera de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales.[64]

La forma jurídicamente correcta de abordar una modificación sustancial de esta naturaleza -que entendemos injustificable- debiera haber sido a través de la reforma de la propia Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General, mediante un debate amplio y cuidadoso reforzando singularmente la publicidad de la reforma, así como respetando la función de los órganos que tienen por misión informar jurídicamente de la compatibilidad de la medida con la Constitución y con las normas de protección de Datos. Acontece que el preceptivo dictamen del Consejo de Estado con número de expediente 757/2017 no aborda en ningún sentido la enmienda de la LOREG debido a que informa del anteproyecto de la Ley Orgánica de Protección de Datos, es decir, informa en una fase previa del proyecto normativo, pero en el *iter legislativo* y en el trámite de enmienda del Senado es donde se introduce esta polémica enmienda de la LOREG. Pensamos que lo anterior es grave porque es una forma de eludir el control de un órgano de relevancia constitucional cuyo origen se debe buscar en la propia redacción de la LO del Consejo de Estado 3/1980, de 22 de abril, y que para evitar estas graves elusiones a su control, debería de *lege ferenda* o bien efectuar su dictamen al final del *iter legislativo*, lo que afectaría negativamente a su eficacia de control *ex ante* y, en todo caso, antes de la promulgación de la norma, o contemplar un doble dictamen previo (proyecto de Ley) y posterior (Ley concluida) para evitar las situaciones como la señalada de dudosa constitucionalidad si atendemos a la función del Consejo de Estado que, como precisa el art 3[65] de su reglamento, es velar por la constitucionalidad,

[64] Dos. Se añade un nuevo artículo cincuenta y ocho bis, con el contenido siguiente: *Utilización de medios tecnológicos y datos personales en las actividades electorales.*

1. *La recopilación de datos personales relativos a las opiniones políticas de las personas que lleven a cabo los partidos políticos en el marco de sus actividades electorales se encontrará amparada en el interés público únicamente cuando se ofrezcan garantías adecuadas.*

2. *Los partidos políticos, coaliciones y agrupaciones electorales podrán utilizar datos personales obtenidos en páginas web y otras fuentes de acceso público para la realización de actividades políticas durante el periodo electoral.*

3. *El envío de propaganda electoral por medios electrónicos o sistemas de mensajería y la contratación de propaganda electoral en redes sociales o medios equivalentes no tendrán la consideración de actividad o comunicación comercial.*

4. *Las actividades divulgativas anteriormente referidas identificarán de modo destacado su naturaleza electoral. 5. Se facilitará al destinatario un modo sencillo y gratuito de ejercicio del derecho de oposición.*

[65] Real Decreto 1674/1980, de 18 de julio, por el que se aprueba el Reglamento Orgánico del Consejo de Estado.

Artículo 3. Constitucionalidad, legalidad y oportunidad.1. En el ejercicio de sus funciones, el Consejo de Estado velará por la observancia de la Constitución y del resto del ordenamiento jurídico. 2. El Consejo de Estado apreciará la legalidad y, en su caso, la constitucionalidad de los

legalidad y oportunidad de las normas sujetas a dictamen. Si se elude dictaminar sobre las mismas porque se introducen a propósito como enmiendas en la fase final del *íter parlamentario* se frustra mediante tal elusión el control que está llamado a ofrecer el Consejo de Estado. Lo mismo ha sucedido mutatis mutandis con la función de control de la Agencia de Protección de Datos la cual no ha podido informar la modificación de la LOREG por la misma razón que hemos señalado en relación con el Consejo de Estado, argumento éste que la propia Agencia de Protección de Datos lamenta en su informe 210070/2018 que se evacúa de forma urgente para ofrecer una respuesta jurídica en materia de protección de Datos precisamente sobre el nuevo artículo 58 bis de la LOREG.

7. ¿PUEDE SER EL INTERÉS PÚBLICO BASE SUFICIENTE PARA LIMITAR SEVERAMENTE VALORES, PRINCIPIOS Y DERECHOS FUNDAMENTALES?

Parece claro que amparar el uso de datos personales de naturaleza política por parte de los partidos políticos en el *interés público* es una estrategia que podríamos denominar de "taimada". Es algo que prohíbe Reglamento 2016/679 Del Parlamento Europeo y Del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE a pesar de lo señalado por el desafortunado considerando número 56 del citado Reglamento que carece de base jurídica, es decir, normativa para desarrollar, a partir de él, consecuencias jurídicas positivas en el marco del derecho interno de los Estados miembros de la UE ya que su naturaleza es meramente la de motivar de modo conciso las disposiciones esenciales de la parte dispositiva (Pascua Mateo: 2006)[66]. Hay que precisar que en la parte dispositiva del reglamente no existe previsión normativa de lo que se argumenta en la motivación del Reglamento, en otras palabras, no existe norma reglamentaria. Sin perjuicio de que su redacción tampoco es congruente con el acervo y tradición comunitaria en materia interpretativa en lo que respecta a la protección de datos como es de fácil deducción de los muy diversos dictámenes y opiniones del extinto grupo de trabajo del artículo 29. Pero el uso de conceptos jurídicos indeterminados descontextualizados, como es el "*interés público*" hace posibles maniobras como la que recoge finalmente nuestra Ley. El problema se encuentra, también aquí, en que no se advierte por ninguna parte el "*interés*

proyectos de disposiciones generales, tratados y actos administrativos sometidos a su consulta y valorará los aspectos de oportunidad y conveniencia cuando lo solicite expresamente la autoridad consultante o cuando lo exija la índole del asunto o la mayor eficacia de la Administración en el cumplimiento de sus fines.

[66] Pascua Mateo, Fabio, "*La técnica normativa en el sistema jurídico comunitario*", Cuadernos de Derecho Público, núm. 28, mayo-agosto 2006, págs. 148-152.

público" en que los partidos políticos -*organizaciones privadas revestidas de funciones públicas*- accedan a ese tipo específico de dato personal político. Simplemente no lo precisan y aunque se compartiese una opinión evolutiva acerca de que los partidos fuesen alguna forma de órganos del Estado *in fieri* y de *lege ferenda*, lo que jurídicamente no son, tampoco se podría comprender ya que las posibilidades de ingeniería social que hace posible esta concepción son radicalmente contradictorias con el respeto de la dignidad de las personas proclamado en el artículo 10 de la Constitución.

La ingeniería de datos pretende tratar a las personas como "*cosas*" los sistemas de propaganda cognitiva virtual tienen por objetivo central *cosificar al elector* explotando su privacidad en favor de la propaganda política partidista.

Que la cosificación haya sido y sea una forma de actuar de los poderes privados en la búsqueda del lucro a través del comercio es muy distinto de que los ciudadanos entren, contra su voluntad, en un mercado en el que la mercancía son precisamente los ciudadanos. Mercancía que, además, no es respetada desde que se cosifica y se explota sus sesgos psicológicos íntimos (lo que incluye patologías médicas) expresados en sus hábitos y conductas de navegación en las redes sociales. El "*interés público*" tiene sentido como concepto restrictivo de derechos, en categorías de datos como los datos de salud, por ejemplo, cuando una pandemia grave puede afectar a la vida de millones de personas, en esos y en muy limitados casos más, el bien superior, la vida, puede ser ponderada por encima de la privacidad individual, sin perjuicio de que en estos casos se adopten un conjunto de garantías que, esas sí, están debidamente protegidas en el Reglamento Europeo de Protección de Datos. Pero, bajo ningún concepto se puede extralimitar el "*interés público*" como instrumento al servicio del interés privado de los partidos políticos, porque no se supera el test de ponderación[67] entre privacidad e intereses privados de los partidos políticos o de poder que éstos representan (Barak, 2017),[68] tampoco se superaría si fuese el Estado quien pudiese utilizar esos datos para funciones electorales. En efecto, en buena medida el uso de los datos personales políticos de los ciudadanos por los partidos políticos es una erosión muy severa de la privacidad, una vuelta al pasado en la evolución de los derechos, es la deconstrucción o la involución de la idea de privacidad acuñada por (Warren y Brandeis: 1995)[69] es una forma de conferir un poder de control injustificado e inmenso de los electores por los partidos políticos que carece de justificación alguna más allá de incrementar la capacidad

[67] Con los tres subprincipios de, idoneidad, necesidad y proporcionalidad en sentido estricto. Es claro que la estructura de fondo de la idoneidad identifica una relación fines-medios.

[68] Barak, Aharon, "*Proporcionalidad. Los derechos fundamentales y sus restricciones*", Palestra, Lima, 2017, págs.385-406.

[69] Warren Samuel y Louis Brandeis, "*El derecho a la intimidad*", Civitas, Madrid, 1995.

de control de estos últimos sobre la sociedad política, lo que es a nuestro juicio inadmisible, parafraseando a (Dworkin: 2018)[70] con ésta regulación no se podrá escapar a su influencia, pero debemos resistir la dominación.

En el sentido señalado las conclusiones del Supervisor Europeo de Protección de Datos, Giovanni Buttarelli sobre la manipulación en línea y los datos personales[71] son a nuestro juicio correctas y plenamente compartibles: "La manipulación en línea supone una amenaza para la sociedad porque las burbujas de filtro y las comunidades cerradas hacen que sea más difícil para las personas entenderse entre sí y compartir experiencias. El debilitamiento de este «pegamento social» puede socavar la democracia, así como varios derechos y libertades fundamentales. La manipulación en línea es también un síntoma de la opacidad y la falta de rendición de cuentas en el ecosistema digital. El problema es real y urgente, y seguramente empeorará a medida que se conecten más personas y cosas a internet y aumente la importancia de los sistemas de inteligencia artificial. En la raíz del problema está en parte el uso irresponsable, ilegal o inmoral de los datos personales. La transparencia es necesaria, pero no suficiente. La gestión de contenidos puede ser necesaria, pero no se puede permitir que comprometa derechos fundamentales. Por tanto, parte de la solución radica en hacer cumplir las normas ya existentes, en particular el RGPD, rigurosa y conjuntamente con otras normas aplicables a las elecciones y al pluralismo en los medios de comunicación".

La explotación del miedo a través de las deepfakes diseñadas con esa expresa finalidad -como vimos más arriba-, de la ira, de las emociones como han estudiado (Parker e Isbell: 2010)[72] y la provocación artificial o inducción de los estados emocionales con base en imágenes -*video*- como analiza (Elinor y colaboradores: 2019)[73] sobre sujetos específicos o grupos concretos a través de circuitos de refuerzo basados en las deepfakes es, precisamente, a lo que puede conducir el uso de los datos personales políticos de los ciudadanos por los partidos, produciendo severas rupturas o distorsiones

[70] Dworkin, Ronald, "*Justicia para erizos*", FCE, México, 2014, pág. 263.

[71] Resumen del Dictamen del SEPD sobre la manipulación en línea y los datos personales, DOUE C 233/8 de 4.7.2018, pág. 11.

[72] Parker, Michael T y Linda M. Isbell, "How I Vote Depends on How I Feel: The Differential Impact of Anger and Fear on Political Information Processing", *Psychological Science* 21(4) 548 - 550.

[73] Elinor Amit, SoYon Rim,Georg Halbeisen, Uriel Cohen Priva, Elena Stephan, Yaacov Trope, "Distance-dependent memory for pictures and words", *Journal of Memory and Language*, N 105, 2019, págs. 119-130.

del espacio común o público como señalan (Marchal, Neudert y colaboradores: 2018)[74] en los electorados objetivos. Una forma posible de radicalizar las campañas electorales y crear divisiones sociales artificiales mediante la polarización de mayor profundidad que las ya existentes en la tradición no virtual. No debemos olvidar aquella admonición que señalara (Duverguer: 1967)[75]de que, en las sociedades desarrolladas la publicidad es el opio del pueblo. Cuando escribía Duverguer no podía imaginar el nivel de refinamiento que la publicidad podía alcanzar y en el campo de la política cognitiva virtual esa advertencia es aún más importante que en el mercado ya que la apelación a las emociones, a los sentimientos tiene una relevancia realmente muy significativa, como ha estudiado (Brader: 2005)[76]en la adopción y cambio de actitudes de los electores, conjuntamente con la manipulación en las formas de generar ansiedad, enfado o esperanzas como señalan (Valentino y colaboradores: 2008)[77]las cuales han tenido un impacto significativo sobre los ciudadanos o electores lo que conduce a fragmentación de la sociedad en su conjunto y a su fragmentación rompiendo valores identitarios comunes como han demostrado (Benkler, Faris y Roberts: 2018)[78]y hemos considerado anteriormente. Debemos recordar por último que nuestra mente (Pluviano y Della Sala: 2019)[79] no es reproductiva, sino reconstructiva. Y en esa reconstrucción de la realidad comete inevitablemente errores, ya que se basa en los propios prejuicios y marcos de referencia. Nuestra mente asimila y recuerda los hechos al vincularlos en un marco de referencia, una narrativa que tratamos de mantener coherente. Por ello, desmontar una información, incluso si se revela incorrecta o falsa, deja un vacío en el modelo mental que se ha creado, generando disonancia cognitiva. Y esa laguna o disonancia la toleramos muy mal, por lo que preferimos un modelo de pensamiento falso pero completo (coherente) a uno parcial (incoherente) y que no encaja con nuestros conocimientos previos.

---

[74] Marchal Nahema, Lisa-Maria Neudert, Bence Kollanyi y Philip N. Howard, "Polarization, Partisanship and Junk News Consumption on Social Media During the 2018 US Midterm Elections" *Data Memo 2018.5*. Oxford, UK: Project on Computational.

[75] Duverguer, Maurice, "*La Democracia sin el pueblo*", Ariel, Barcelona, 1967, pág. 262.

[76] Brader, Ted, "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions", *American Journal of Political Science*, Vol. 49, No. 2. (Apr, 2005), págs. 388-405.

[77] Valentino, Nicholas A, Vincent L. Hutchings, Antoine J. Banks, "Is a Worried Citizen a Good Citizen? Emotions, Political Information Seeking, and Learning via the Internet", *Political Psychology,* Vol. 29, No. 2, 2008.

[78] Benkler, Yochai, Robert Faris, y Hal Roberts, "*Network Propaganda, Manipulation, Disinformation, and Radicalization in American Politics*", Oxford University Press, 2018.

[79] Pluviano, Sara, Sergio Della Sala, "El autoengaño de los antivacunas", Op Cit, pág. 22.

Estos aspectos no los desconoce la Comisión Europea y así lo ha reflejado en su documento de orientación titulado "*Elecciones libres y limpias*"[80] cuando señala: El Comité consideró que la publicidad personalizada en línea podría ser capaz, en algunas circunstancias, de afectar significativamente a las personas cuando, por ejemplo, es intrusiva o aprovecha su conocimiento de aspectos vulnerables de las personas. Dada la importancia del ejercicio del derecho democrático al voto, los mensajes personalizados cuyo posible efecto sea, por ejemplo, que las personas no voten, o voten de una forma específica, podrían potencialmente cumplir el criterio de efecto significativo. Por tanto, en el contexto electoral, los responsables del tratamiento deben garantizar que todo tratamiento que utilice esas técnicas es lícito con arreglo a los citados principios y las estrictas condiciones expuestas en el Reglamento general de protección de datos"

Pese a lo señalado el Reglamento Europeo de Protección de Datos facilita estos tratamientos. Muestra de lo anterior en nuestra nueva regulación normativa es la expresión contenida en el artículo 58 bis de la LOREG. Recordemos que en el número 1 sobre la Recopilación de datos relativos a las opiniones políticas de las personas, ese número concluye con la expresión: "*únicamente cuando se ofrezcan garantías adecuadas*". Las garantías no se determinan ni enumeran en la propia norma, lo que sería lo jurídicamente debido, tan sólo se señala la expresión vacía "garantías adecuadas" remitiendo su determinación quizá a una norma de desarrollo reglamentario. La propia Agencia de protección de datos se sorprende de que un elemento tan importante como las garantías no queden concretamente especificadas en la misma Ley Orgánica.[81]

Al margen de lo anterior, de por si grave, la Agencia de Protección de Datos en un informe jurídico remitido por su propio Gabinete Jurídico, referente a la consulta planteada por la directora de la AEPD sobre el tratamiento de datos relativos a opiniones

---

[80] "Elecciones libres y limpias", Orientaciones de la Comisión relativas a la aplicación de la legislación sobre protección de datos de la Unión en el contexto electoral, Contribución de la Comisión Europea a la reunión de dirigentes en Salzburgo los días 19 y 20 de septiembre de 2018, COM (2018) 638 final, págs. 8-9.

[81] Señala concretamente la AEPD: "Por ello, hubiera sido conveniente que las mismas se establecieran en el texto del propio artículo 58 bis, haciendo uso de la habilitación contenida en el artículo 6.2 y 3 del RGPD, entre las que se podrían haber incluido la relativas a: *"condiciones generales que rigen la licitud del tratamiento por parte del responsable; los tipos de datos objeto de tratamiento; los interesados afectados; las entidades a las que se pueden comunicar datos personales y los fines de tal comunicación; la limitación de la finalidad; los plazos de conservación de los datos, así como las operaciones y los procedimientos del tratamiento, incluidas las medidas para garantizar un tratamiento lícito y equitativo, como las relativas a otras situaciones específicas de tratamiento a tenor del capítulo IX. El Derecho de la Unión o de los Estados miembros cumplirá un objetivo de interés público y será proporcional al fin legítimo perseguido"*. Asimismo, dicha ley podría haber establecido la obligación de consultar a la autoridad de control y de recabar su autorización previa en relación con el tratamiento por un responsable en el ejercicio de una misión realizada en interés público, conforme al artículo 36.5 del RGPD". Agencia Española de Protección de Datos, informe jurídico Ref210070/2018.

políticas por los partidos políticos al amparo del artículo 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General, trata de fijar las garantías que debiera haber fijado el legislador con rango orgánico. El informe de la AEPD, naturalmente, carece de rango normativo para fijar lo que deberían ser las garantías mínimas para realizar un tratamiento que, a nuestro juicio, no es constitucional porque vulnera la dignidad de los ciudadanos contemplada en el artículo 10 de nuestra norma suprema ya que la norma que consideramos tiene directa incidencia sobre el libre desarrollo de la personalidad de los ciudadanos desde el momento en el que ésta puede ser condicionada por partidos políticos que traten de dirigirla a través de tecnologías de propaganda psicológica que explota sus sesgos cognitivos, que los partidos políticos ni tienen el derecho de conocer ni el derecho a usar en contra de la voluntad libre del ciudadano. Así como violenta y sacrifica el apartado segundo del artículo 16 de la Constitución que señala: "*Nadie podrá ser obligado a declarar sobre su ideología, religión o creencias*". El acceso por los partidos políticos a los datos personales políticos de los ciudadanos representa una forma de auto declaración contra la voluntad de los ciudadanos, se trata de una forma de declaración "obligatoria" explícita, ya que se extrae por medio de un algoritmo de minería de datos que analiza la conducta del ciudadano en Internet contra la voluntad de sus titulares mediante el uso de instrumentos de procesamiento de información Big Data, lo que es y debería seguir siendo una conducta privada que debe estar preservada de la mirada curiosa y atenta tanto de poderes privados como de los poderes públicos con finalidades políticas, es decir, de poder. El 22 de mayo de 2019 el Pleno del Tribunal Constitucional por unanimidad ha declarado contrario a la Constitución y nulo el apartado 1 del art. 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del Régimen Electoral General, que permite a los partidos políticos recoger datos personales relativos a las opiniones políticas de los ciudadanos. La sentencia, cuyo ponente ha sido el Magistrado Cándido Conde-Pumpido, ha estimado el recurso de inconstitucionalidad presentado por el Defensor del Pueblo el pasado 5 de marzo de 2019.

8. LA RUPTURA DEL ESPACIO PÚBLICO COMÚN Y LA POLARIZACIÓN PUEDEN CONDUCIR A LA AUTOCENSURA

Es claro, en conexión con lo anterior, que toda forma de control de la actividad de los ciudadanos en las redes supone una restricción de mayor o menor intensidad sobre la libertad de expresión, art. 20 de la Constitución tanto en su vertiente activa, emisión de pensamientos como sobre la pasiva, recepción de estos.

No podemos dejar de recordar por ello las certeras y siempre actuales consideraciones de John (Stuart Mill: 1997)[82] para quien: "El valor de un estado, a la larga, es el valor de los individuos que lo componen; y un estado que pospone los intereses de la expansión y elevación mental de sus individuos, a un poco más de perfección administrativa o a la apariencia que de ella da la práctica en los detalles de los asuntos; un estado que empequeñece a sus hombres, a fin de que puedan ser más dóciles instrumentos en sus manos, aun cuando sea para fines beneficiosos, hallará que con hombres pequeños ninguna cosa grande puede ser realizada; y que la perfección del mecanismo, a la cual todo lo ha sacrificado, terminará por no servirle para nada por falta de poder vital que, en aras de un más fácil funcionamiento de la máquina, ha preferido proscribir". En efecto, un intercambio de ideas carente de inhibiciones (basadas en sutiles y menos sutiles formas de censura) y de información robusto (la expansión y elevación mental que preconiza Mill), en el sentido de autónomo con respecto a los medios de comunicación de masas convencionales, puede predisponer y formar ciudadanos no sólo bien informados, sino, y lo que es más importante, "formados", amén de fortalecer las relaciones comunicativas ciudadanas, lo que no supone otra cosa que construir una estructura democrática para el diálogo (vocación integradora del medio de comunicaciones) y una ciudadanía, a la postre, mejor formada y capacitada para participar en los procesos democráticos, según los modelos institucionales informales o formales que se puedan articular. Internet, no puede dudarse, constituye una fuente inagotable de expresión, de intercambio de ideas. Como ya advirtiera (Milton: 2000)[83] con ocasión de la respuesta ante la orden del Parlamento Británico, de 14 de junio de 1643, por la cual se redacta la Aeropagítica, el autor señalaba: "Donde hay mucho deseo de aprender es natural que haya mucho que argumentar, muchos escritos, diversas opiniones: porque la opinión en la gente de valía no es más que un conocimiento en desarrollo y formación. Con estos fantasiosos pavores a cismas (en referencia a la orden que autorizaba la censura previa) y sectas, estamos perjudicando el serio y celoso afán de conocimiento y de comprensión que Dios ha despertado en la gente de esta ciudad..." La confianza de que las comunicaciones se están desarrollando de un modo privado y fiable, fomenta la razonable esperanza de que el medio de comunicación responde a las expectativas de privacidad que de él son exigibles y conduce a un uso del instrumento de comunicaciones adecuado a las necesidades de cada usuario, sin que el temor fomente en cada interlocutor un proceso de autocensura en las expresiones empleadas o en las informaciones que se pretendan comunicar, que vaciaría de contenido el

[82] Stuart Mill, John, *Sobre la libertad*, Alianza Editorial, 1997.

[83] John Milton, *Aeropagítica,* Ed. Torre de Goyanes, 2000, pág. 103.

concepto de una opinión pública libre en este medio de telecomunicaciones, donde la libertad encuentra su garantía de que no es condicionada por intercepciones ilegítimas o usos espurios de los datos personales que coaccionen la calidad, variedad, tipo o naturaleza de cualquier forma de expresión a la que el medio pueda conferir operatividad. Es decir, induce en la comunidad política la percepción correcta de que una injerencia irrestricta se ejerce legalmente sobre sus comunicaciones personales y que, conduce a limitar el ejercicio de los derechos fundamentales, no sólo en lo que respecta al contenido de lo comunicado, sino igualmente en el proceso *no condicionado y libre de búsqueda de información*, faceta comprendida en el derecho de opinión y expresión del artículo 19 de la Declaración Universal de Derechos Humanos, de 10 de diciembre de 1948, actividad en la actualidad vigilada y productora, razonablemente, de temor, generador de autocensura. En estas condiciones de vigilancia masiva de todos los ciudadanos (que hagan uso de Internet en España), la autocensura, tal vez la forma de censura más dramática y restrictiva de la libertad ya que afecta directamente a la dignidad del hombre en su fuero más interno, daña irremediablemente la vertiente subjetiva de la persona, como erosiona simultáneamente el derecho reactivo de defensa frente al Estado que la libertad de expresión contiene en su seno y, objetivamente considerada, destruye la institución política de la opinión pública libre -en este medio de comunicación-, dado que nutrir un instituto fundamental para la democracia de afluentes debilitados, de expresiones sin vigor y atemorizadas, es generar un plasma insustancial, estéril e inútil que escasos servicios puede rendir a la democracia, debido esa autocensura que produce el denominado "*Chilling Effect*" o efecto de enfriado de la expresión; el retraimiento de los hablantes o la autorestricción en la expresión por el temor al castigo desproporcionado.

Hay que ser conscientes de que la norma no ha sido elaborada por los ciudadanos, objeto de esta, sino precisamente por aquellos, los políticos profesionales que encuentran en las tecnologías de Big Data y procesamiento de la información que consideramos, un instrumento como jamás habían podido imaginar tener en sus manos para controlar o, al menos intentarlo a sus electores. Por ello es poco realista pensar que aquellos que se van a beneficiar de su inmenso poder pongan reparos a su aparición y desarrollo, el razonamiento alcanza igualmente a las instituciones Europeas que han desarrollado un Reglamento de Protección de Datos que mediante conceptos jurídicos indeterminados, pero fácilmente determinables por quienes se benefician de ellos, no han tenido reparos en adoptar un Reglamento que, en la materia que nos ocupa, es en alguna medida responsable de lo sucedido en la regulación española. Los Europarlamentarios y los miembros de la Comisión forman parte de los políticos profesionales que produce la democracia representativa de partidos. Y, si bien, se

realizan admoniciones contra el poder de los nuevos instrumentos técnicos al servicio de la política, la realidad es que pudiendo haberlos limitado de forma efectiva esto no se ha hecho. Detrás de ello parece encontrarse una concepción *transpersonalista* de los derechos que ignora lo que a nuestro juicio sería la opción coherente con la tradición europea respetuosa de los derechos fundamenta que sería la *personalista* y que es la que debe ser restaurada tras esta agresión. Es decir, y con Vanossi[84] en mantener y defender la concepción *personalista*, de raigambre Kantiana que, según Radbruch, determina que consideremos a la persona humana como un fin en sí mismo, en lugar de admitir el sentido opuesto de las concepciones *transpersonalistas*, para las que el Estado es fin en sí mismo, mientras que el hombre es tan sólo un medio. El hombre no puede ser utilizado únicamente como medio por ningún hombre, ni por otros, ni siquiera por sí mismo, sino siempre a la vez como fin y en eso consiste, precisamente, su dignidad (la personalidad), elevándose sobre *las cosas*[85]. Se cosifica a un ser humano cuando, olvidando o desconociendo su dignidad, se le presenta a los demás miembros de la comunidad como "*cosa*" que es precisamente lo que permiten las campañas electorales cognitivas virtuales, cosificar al elector mediante la justificación de que existe un interés público en que ello sea así para fortalecer el ya inmenso poder de los partidos políticos como si se tratasen de órganos del Estado, no que ejerzan *funciones de órgano del Estado* como recordara (Leibholz: 1980)[86] citando la célebre decisión de Pleno del Tribunal Constitucional Federal alemán de 20 de julio de 1954, y degradando en vez de fortaleciendo una democracia ya de por si deteriorada.

9. LAS DEEPFAKES SE DIRIGEN A EXPLOTAR LAS EMOCIONES

Como argumentan (Bauman y Donskis: 2019)[87]la transición de lo kafkiano a lo orwelliano marca la línea divisoria entre el mal sólido y el mal líquido, un mal liquido difuso, adaptable y que rellena o puede rellenar todas las cavidades otrora reservadas a espacios de libertad no anegados por ese control líquido. El ciudadano de cristal, aquella distopia que se creía imposible y a la que no se podía llegar, finalmente está llegando y se está asentando en nuestras sociedades, las normas de protección de datos no garantizar de forma eficiente la protección adecuada y necesaria de los ciudadanos y de su privacidad en las redes sociales y en sus interacciones de navegación. Todo esto

[84] Jorge Reinaldo A. Vanossi, "*El Estado de Derecho en el constitucionalismo social*", Eudeba, Argentina, 3ª Edición, 2000, pág. 26.

[85] Inmanuel Kant, "*La metafísica de las costumbres*", Tecnos, Madrid, 1994, pág. 335.

[86] Leibholz, Gerhard, "Representación e identidad", en: Kurt Lenk y Franz Newmann (eds.), "*Teoría y sociología críticas de los partidos políticos*", Anagrama, Barcelona, 1980, pág. 208.

[87] Bauman, Zygmunt y Leonidas Donskis, "*Maldad liquida*", Paidós, Barcelona, 2019, pág. 131.

conduce a nuevas formas de control social que nos alejan del ciudadano responsable y dueño de márgenes de libertad aceptables para transformarlo en una especie de súbdito digital que no es capaz de autodeterminarse en el entorno virtual y al que se le analiza como un espécimen bajo la lente del microscopio para saber qué piensa, qué siente, que desea y sobre esos datos formular tentativas para cambiar, orientar y conducir su vida.

El control de la sociedad y de la conducta individual es lo que hemos venido examinando, de aquí a sistemas de puntuación ciudadana o "*crédito social*"[88] como el que ya se ensaya en China con su proyecto totalitario "*Internet Plus*"[89] hay poca distancia si no se adoptan medidas vigorosas que debiliten de forma consistente la capacidad de control y manipulación de los datos por organizaciones, Estados he intereses públicos y privados. Esa puntuación ciudadana ya es operativa en la actualidad y ya está entre nosotros, por ejemplo, para la fijación de las condiciones en que se puede obtener un crédito, acceder a un empleo o quizá contratar una póliza de seguro médico. Pero también puede generar un neo paternalismo del Estado en el que el Gobierno no sólo esté interesado en lo que hacemos, sino que también se asegure de que hacemos lo que debemos, como preparar por nosotros nuestras declaraciones de impuestos. Creemos que, una fórmula adecuada para cambiar el paradigma es permitir que la información tenga valor para quien la genera, es decir patrimonializar[90] los datos de carácter personal y que estos, sean de titularidad de los ciudadanos, argumento defendido entre otros por (Pentland: 2018).[91] (Hoffmann-Riem: 2018)[92] señala -sin abordar si pudiese y en qué medida existir un derecho exclusivo o comparable con o idéntico a la propiedad- que, desde un enfoque jurídico-político sería incluso defendible, cuando no imperioso, obligar a los proveedores a pagar a los usuarios una remuneración justa cuando consientan la colecta y utilización de datos especialmente valiosos para el tratamiento. Las opciones que se abren a las sociedades democráticas no están cerradas completamente, pero se están cerrando, el momento de tomar decisiones sigue siendo posible, pero se avecinan tiempos complicados precisamente con la Internet de los objetos que estrecharán el ya

[88] El programa del gobierno Chino puede verse en: https://chinacopyrightandmedia.wordpress.com/2014/06/14/planning-outline-for-the-construction-of-a-social-credit-system-2014-2020/ (consultado 12/05/2019)

[89] https://www.independent.co.uk/news/world/asia/china-surveillance-big-data-score-censorship-a7375221.html (consultado 12/05/2019)

[90] González de la Garza, Luis Miguel, "¿Patrimonializar los datos de carácter personal? Argumentos para un debate", *El Notario del Siglo XXI*, nº 38, 5 de julio de 2011. http://www.elnotario.es/index.php/hemeroteca/revista-38/743-patrimonializar-los-datos-de-caracter-personal-argumentos-para-un-debate-0-022018592825176746

[91] Pentland, Alex, "*Una sociedad dirigida por datos*", Investigación y Ciencia, nº 91, 2018, pág. 17.

[92] Hoffmann-Riem Wolfgand, "*Big Data. Desafíos también para el Derecho*", Civitas, Navarra, 2017, pág.116.

estrecho círculo de lo que se podrá saber de cada ciudadano sin ventaja alguna para él. Algo análogo sucede con la tecnología 5G en la que China y los Estados Unidos pujan por definir estándares ya que la definición del continente determinará el contenido y las formas de su apropiación tecnológica desde la perspectiva del derecho de patentes.

La libertad de expresión corre igualmente riesgos ciertos. Cuando el ciudadano pensaba y sabía que lo que hacía era una actividad que no estaba siendo monitorizada por el escrutinio de organizaciones privadas y públicas, podía navegar con libertad por donde le llevase su curiosidad en el ejercicio de su libertad por saber, por conocer, en suma, por aprender. Cuando el ciudadano empieza a saber que todo movimiento es almacenado, registrado y analizado para extraer cuidadosamente consecuencias analíticas de su conducta, la libertad puede dejar paso a la restricción por el temor cierto de que un conjunto de algoritmos anónimos y con una lógica secreta -en el sentido de que los algoritmos no funcionan libres de errores o de una programación dirigida y finalista, con lo que los principios de equidad y justicia se verán reemplazados por la arbitrariedad de la *Ley de la programación*, Ley contra la que los ciudadanos no podrán defenderse-, lo clasifiquen y, en virtud de esa clasificación, se puedan sufrir perjuicios derivados de una simple e inocua forma de ser o de ejercer la libertad. Si se escribe, si se opina, si se discute o se visita una página web sospechosa, es perfectamente posible recibir una etiqueta electrónica clasificatoria en algún lugar geográfico del vasto mundo virtual -posiblemente no sujeto a la jurisdicción Europea- en lo que damos en llamar el cambio de paradigma de una sociedad de clases a una sociedad clasificada, donde el ciudadano pueda formar parte de índices de buena o mala conducta ciudadana o social, de ser un activista y mal ciudadano sospechoso o alguien con una conducta ordenada según los estándares que se crean apropiados por el mercado o por el Estado. (Helbing, Freig y colaboradores: 2018)[93] sostienen y es un argumento que compartimos que, para crear transparencia y confianza suficientes las principales instituciones científicas deberían actuar como depositarias de la información y de los algoritmos, que actualmente escapan al control democrático. Ello también requeriría un código de conducta adecuado que, como mínimo, deberían cumplir todos los que tuviesen acceso a la información y algoritmos delicados: una suerte de "*juramento hipocrático*" para los profesionales de la tecnología de la información,

La libertad de expresión genera ideas útiles para la sociedad ideas que difundidas y recombinadas se pueden transformar en innovación, la fragmentación de las redes y el

[93] Helbing Dirk, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari y Anfrej Zwitter, "Una estrategia para la era digital. Una hoja de ruta para evitar el retroceso de la democracia en la sociedad de la información", *Investigación y Ciencia, Temas*, "La Sociedad hiperconetada", nº 91, 2018, págs.74-75,

temor a actuar con libertad mediante la autocensura -sin desconocer la responsabilidad de los propios actos- pueden conducir a una era de repliegue de las ideas, encerradas en esferas, en redes aisladas unas de otras con pérdida de libertad general para la actividad de los ciudadanos en la sociedad en su conjunto transformándose las redes en un sistema de control como nunca antes se había pensado disponer, sin desconocer como ya señalamos que las redes no son en absoluto homogéneas como no lo es la sociedad. Nuestra experiencia actual es tributaria del posmodernismo que ha pretendido diluir en un "relato fragmentado" la realidad y en el que la "verdad" queda reducida a la "retórica victoriosa" de las ideologías triunfantes en un momento histórico dado, podríamos decir que es verdadero aquello que consigue imponerse como tal sin tener en cuenta si existe o no base alguna para ello. Para esta visión equivocada pero muy real el "pasado" es una "narrativa" cuyo valor de verdad se resuelve íntegramente en el presente, siendo el presente el que se "inventa" ese pasado en función de intereses políticos o ideológicos. La ignorancia, el miedo y la confusión conducen a nuevas formas de "*conformidad*" social y pasividad por indiferencia a la verdad, en alguna forma ya advertidas hace siglos, por ejemplo, por Etienne de la Boétie en su no superado discurso sobre la Servidumbre voluntaria. Pero esa conformidad puede traducirse en muy altos costes sociales, es decir, pensemos por ejemplo en el fenómeno de los movimientos antivacunas, padres y madres que por ignorancia desconocen la importancia de estas para sus hijos y para la salud de la sociedad en su conjunto. Es información falsa que alimenta a grupos polarizados que han creado asociaciones basadas en el error, como aquellas otras que piensan que la tierra es plana, la cuestión son los costes sociales de unas ideas con respecto a otras y si es legítima la acción del Estado para proteger a las victimas de esa información falsa. Como señalara (Laski: 2017)[94] no hay duda de que el precio de la inercia es, a la larga la pérdida del sentido cívico entre la multitud. Cuando se erosiona sistemáticamente la verdad la sociedad no reacciona en su conjunto cuando observa la proliferación de este tipo de movimientos de naturaleza sectaria y basados en una ignorancia agresiva o militante y se recluye dudando de sus propios conocimientos y temiendo en muchos casos hacer pública su opinión.

10. UN NUEVO DERECHO *"EL DERECHO A NO SER ENGAÑADO"*

¿Por qué surge la necesidad de articular un nuevo derecho en la sociedad global de la información? Seguramente porque no existe una fórmula en la que no dando entrada al Derecho se articule mejor la garantía de que las personas no sean tratadas como "*cosas*" a las que se les puede engañar cuando sea del interés particular de los

[94] Laski, Harold, "*Los peligros de la desobediencia*", Sequitur, Madrid, 2017, pág. 32.

ciudadanos que forman parte profesional o accidental de lo que denominados la política tanto nacional como internacional o se sea el centro de atención de campañas de desinformación de organizaciones privadas o públicas con las más variadas finalidades que podamos imaginar.

Engañar a las personas es usarlas como instrumentos, es despreciar su dignidad y menospreciar el respeto que unos nos debemos a otros en el marco de una sociedad cooperativa y organizada (Kant: 2012)[95]. La mentira puede ser estudiada de forma útil como *un costo en las transacciones informativas humanas* que las hace dificultosas o imposibles. Creemos de interés clarificar la definición de "*costo de transacción*" con (Williamson: 2009)[96]. Así, ocurre una transacción cuando se transfiere un bien, servicio o una *expresión* a través de una interfase tecnológicamente separable. Termina una etapa de la actividad y se inicia otra. Con una interfase que funcione bien, como en el caso de una máquina que funcione bien, estas transferencias ocurren suavemente. En los sistemas mecánicos, buscamos eliminar las fricciones: ¿encajan los engranajes, están lubricadas las piezas, hay fugas innecesarias u otras pérdidas de energía? La contraparte económica y organizativa -añadimos nosotros- es el costo de transacción: ¿operan armoniosamente las partes de la transacción o hay frecuentes malentendidos y conflictos que generan demoras, descomposturas y otras deficiencias del funcionamiento? La mentira puede pues verse como una pieza que no encaja armónicamente con la *realidad*, una disfunción de la pieza informacional que se diferencia de la verdad precisamente en su falsedad y que por ello genera conductas en las diversas etapas del uso de la información erróneas, equivocadas y en los peores casos situaciones personales o sociales desastrosas ya que se han adoptado con base en esa información (Holbach: 2016)[97] lo que supone incidir en el honor de quien ha actuado en base a esa información tendenciosamente falsa. La articulación de un derecho a la verdad es de suma complejidad porque determinar y definir qué es la verdad en múltiples dimensiones es una tarea hercúlea e imposible, en cambio, determinar un derecho a no ser engañado invierte la carga de la prueba en multitud de dimensiones en las que el titular del derecho *a no ser engañado* puede exigir una investigación imparcial del poder público en el caso concreto y, de determinarse el engaño, establecer su reparación, no hacemos aquí referencia al engaño característico de los tipos penales como en el delito de estafa del art 248 de la LO 10/1995, de 23 de noviembre, del Código Penal. Quizá la sede normativa de este derecho *a no ser*

[95] Kant, Inmanuel, Benjamin Constant, *"¿Hay derecho a mentir?"*, Tecnos, Madrid, 2012, pág. 29.

[96] Williamson, Oliver E., "*Las instituciones económicas del capitalismo*", FCE, México, 2009, págs. 13-18.

[97] Holbach, "*Ensayos sobre los prejuicios*", Laetoli, Navarra, 2016, págs. 34-35.

engañado debería ocupar el lugar de la vetusta LO 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen. Con un alcance actual en el que el derecho al honor pudiese subsumirse en un nuevo derecho *a no ser engañado*.

En el caso que hemos venido considerando, las mentiras profundas de base audiovisual, su capacidad de engañar y por ello de producir efectos sociales realmente disruptivos es enorme, porque permiten también manipular el pasado. Podemos pensar, por ejemplo, en diseñar a Kennedy diciendo cosas que jamás dijo, o a Lenin o a quien deseemos formular relatos falaces. Se genera su imagen, se genera su voz, pero haciendo o diciendo lo que nunca dijo, con una finalidad política concreta. Eso es manipular la historia, recrear hechos que nunca pasaron. Pero la clave es ¿puede el ciudadano medio que carece de fuentes veraces discernir si esa información es real o no? La respuesta en muchos casos será "no". Y ahí está la capacidad disruptiva profunda, es decir, la falsificación de la historia mediante el engaño y simultáneamente la erosión de la confianza en la historia y en los relatos veraces caldo de cultivo idóneo para la creación de neopopulismos capaces de poner en duda los relatos facticos que fundamentan nuestra realidad colectiva.

Mediante el derecho a no ser engañado esas imágenes deberían llevar necesariamente una etiqueta o meta etiqueta que señale que ese producto audiovisual no es original o genuino, sino una ficción. De forma que el receptor de esta pueda saber que es una pieza de diseño audiovisual que no es "veraz". Surge el problema ¿Quién debe clasificar las piezas audiovisuales como verdaderas o falsas? El propio creador, la sociedad de forma descentralizada de forma análoga a cómo funciona la Wikipedia. Esa es una buena pregunta. Pensamos que, en un entorno Global, el modelo de la Wikipedia puede ser interesante, un sistema de valoración de las imágenes descentralizado y eficiente gestionado por una comunidad de personas expertas (Polanyi: 2009)[98]. Hoy las imágenes se pueden filtrar y hacer virales por cualquier parte del mundo y seguramente un control global podría ser la forma inicialmente optima de realizar tal control, asimismo las plataformas, especialmente You Tube juegan un papel destacado en el control de la realidad y en el cumplimiento del derecho a no ser engañado. Pero también se hace urgente y completamente necesario -como señalamos más arriba- crear tribunales *on-line* para "*en tiempo real*" solicitar una "orden de cesación" y que esas imágenes sean cautelarmente retiradas por orden judicial de los servidores de almacenamiento hasta que se determine su naturaleza. La justicia electrónica juega aquí un papel muy relevante en conexión con el derecho a no ser engañado. Si entendemos que los

[98] Polanyi Michael, *"La lógica de la libertad"*, Katz, Madrid, 2009, pág. 129 y ss.

derechos fundamentales son límites al poder, y hoy en día el poder se ostenta no sólo por el Estado sino también por los particulares, singularmente por empresas y organizaciones de base privada de fuerza cuasi estatal, resulta razonable expandir la eficacia de estos derechos a las relaciones privadas, según sostiene la teoría de la Drittwirkung. Entendemos que el derecho a no ser engañado tiene que responder a esta estructura dogmática para que sea un derecho eficiente en un nuevo entorno social y político.