IDP REVISTA D'INTERNET, DRET I POLÍTICA
REVISTA DE INTERNET, DERECHO Y POLÍTICA

https://idp.uoc.edu

Deepfake y desinformación –¿Qué puede hacer el derecho frente a las noticias falsas creadas por *deepfake*?

Gergely Ferenc Lendvai
Universidad Católica Pázmány Péter

Gergely Gosztonyi
Universidad Eötvös Loránd (ELTE)

Fecha de presentación: marzo 2024
Fecha de aceptación: mayo 2024
Fecha de publicación: octubre 2024

Resumen

Este artículo explora la relación entre *deepfake* y *fake news* a través de la lente del derecho. Se divide en tres partes conceptuales principales: en primer lugar, se examina la base teórica, a continuación, se explora la relación entre desinformación y *deepfake* y, por último, se ofrece una visión general de la regulación pertinente en el contexto del contenido de noticias falsas deepfake (en particular las tendencias reguladoras de Estados Unidos, Europa y China). El documento presta especial atención a la percepción jurídica de la desinformación en el contexto de la tecnología *deepfake*, destacando las consecuencias sociales y jurídicas perjudiciales que conlleva. Se basa en una visión general de las normativas nacionales e internacionales y de la bibliografía pertinente, e incluye las soluciones propuestas por los autores a las controversias generadas por la desinformación *deepfake*.

Palabras clave

manipulación; regulación de la IA; DSA; *deepfake*; desinformación; libertad de expresión

IDP REVISTA D'INTERNET, DRET I POLÍTICA
REVISTA DE INTERNET, DERECHO Y POLÍTICA

https://idp.uoc.edu

Deepfake and disinformation – What can the law do about fake news created with deepfake?

Abstract

This paper explores the relationship between deepfake and fake news through the lens of the law. It is divided into three main conceptual parts; first, it looks into the conceptual basis, then it explores the relationship between disinformation and deepfake, and finally, it offers an overview of the relevant regulation in the context of deepfake fake news content (in particular, the US, the European and the Chinese regulatory trends). The paper pays particular attention to the legal perception of disinformation in the context of deepfake technology, highlighting the harmful social and legal outcomes involved. It is based on an overview of national and international regulations and of the relevant literature and includes the authors' proposed solutions to the controversies generated by deepfake disinformation.

Keywords

manipulation; AI regulation; DSA; deepfake; disinformation; freedom of expression

Introducción

Viktor Orbán, el primer ministro de Hungría, es notoriamente reacio a la migración hacia Europa y quiere preservar el «carácter cristiano» del continente (Rankin, 2018). Sin embargo, ¿podrías encontrar un vídeo en el que diga que es un gran fan de la migración y que le gustaría derribar todas las fronteras y vallas alrededor de Hungría para acoger a cualquiera que pida ayuda porque ha tenido que abandonar su patria debido a graves crisis? Ahora parece imposible. Pero ¿es realmente imposible? Todo lo que se necesitaría es un programa gratuito, tantas fotos como sea posible, un ordenador potente y tiempo suficiente para hacerlo realidad. Así que la pregunta es: ¿puede deepfake cambiar nuestras vidas? ¿Puede cambiar lo que percibimos con nuestros propios ojos? En este artículo tratamos de responder a la pregunta de cómo puede el Derecho combatir la realidad generada por la inteligencia artificial (IA) y la desinformación.

En primer lugar, es importante subrayar que la tecnología *deepfake* en sí no es –*per se*– «mala» o «prohibible». La tecnología tiene usos increíblemente útiles en el ámbito de la cultura (como la «reencarnación» de Dalí en un museo), en la concienciación (como el anuncio en el que una versión *deepfake* de David Beckham llama a la acción contra la malaria en nueve idiomas) o en la creación de parodias o contenidos humorísticos/entretenidos (como las parodias de Trump en los *late night shows* estadounidenses) (Lendvai, 2023). El tratamiento de los contenidos *deepfake* ilegales o perjudiciales es extremadamente difícil, sobre todo si se tiene en cuenta la gran complejidad jurídica de la desinformación (véase: Espaliú-Berdud, 2022). Aunque la legislación parece ir con retraso, un país, el Reino Unido, ha encontrado ya una respuesta a la cuestión de la criminalización de los *deepfakes*. La Ley de Seguridad en Línea (Online Safety Act) de 2023 impone a los servicios regulados la responsabilidad de impedir la publicación de determinada información falsa, lo que indica un esfuerzo legislativo por abordar los problemas de seguridad en línea (Coe, 2023). Además, la normativa británica es el primer instrumento jurídico que penaliza directamente la pornografía *deepfake*, ya que las disposiciones sobre el abuso de imágenes íntimas (Sección 188) tipifican como delito el intercambio no consentido de *deepfakes*.

En el presente documento pretendemos debatir tres puntos clave:

1) la definición y la ambigüedad de la conceptualización del *deepfake*,

2) las resoluciones y propuestas legales existentes para regular el *deepfake*, y

3) el camino a seguir en la regulación del *deepfake*.

En cuanto al primer punto estructural, pretendemos sentar unas bases globales para entender qué es el *deepfake*.

Esto es fundamental, ya que debería haber un concepto común en torno al cual las normativas divergentes estipulen el fenómeno. En cuanto a las normativas existentes, pretendemos diferenciar tres filosofías reguladoras: la estadounidense, que opta por una regulación más liberal, la europea, que regula en función de los riesgos, y la china, cuyo objetivo es regular la «raíz de la polémica»: los desarrolladores e ingenieros (Hine y Floridi, 2022). Por último, pretendemos proponer respuestas a cuestiones clave que no están debidamente reguladas.

Para investigar estos tres elementos estructurales clave, la investigación se basa en el análisis jurídico comparativo, un valioso método de investigación que pretende identificar las tendencias modernas, las convergencias y las divergencias entre los sistemas jurídicos (Spina, 2024). Este método de investigación es normativo y descriptivo y permite comprender el papel interdisciplinario de la comparación jurídica y del método de análisis de casos/reglamentos, así como la apertura de la comparación a otros conocimientos y los retos de la interdisciplinariedad (Husa, 2022). En este caso, permite una visión holística de la normativa sobre *deepfakes*, centrándose específicamente en los contenidos que pueden entenderse como desinformación.

1. Marco conceptual: ¿qué es el *deepfake*? «Plagio facial» y *deepfake*

El término *deepfake* es el resultado de la combinación de las palabras *deep learning* (aprendizaje profundo) y *fake* (falso); el primer elemento del término hace referencia al aprendizaje profundo, mientras que el segundo elemento se refiere a la naturaleza falsa del contenido producido (Citron y Chesney, 2019). Aunque el concepto de *deepfake* puede abordarse desde varios ángulos, desde una perspectiva jurídica merece la pena utilizar la definición de *deepfake* de múltiples fuentes de Gergely Ferenc Lendvai (2023) para sentar las bases conceptuales. Según este concepto, *deepfake* es una tecnología basada en IA que permite a un usuario crear de forma deliberada e intencionada contenidos de imagen y audio falsos, principalmente de personas, de tal manera que el contenido creado pueda hacer que el contenido falso parezca real de forma convincente.

En cuanto a los detalles tecnológicos, existen innumerables posibilidades de creación de *deepfakes*, siendo la más conocida la tecnología GAN, o *Generative Adversial Networks* (Pantserev, 2020). Esta tecnología es una red neuronal, un generador que crea nuevos contenidos falsos descargando conjuntos de datos y contenidos de datos. Esto significa que, al introducir cierto material de imagen y audio, la tecnología *deepfake* es capaz de mapear el material que proporcionamos en otro contenido diferente, es decir, basándose en imágenes o grabaciones de audio disponibles de una persona, crea una grabación manipulada que muestra a la persona real en una situación ficticia en la que la persona específica no estaba, o le atribuye una declaración que no hizo.

Entonces, ¿qué convierte al *deepfake content* en un asunto de desinformación? La piedra angular de nuestro argumento es que el *deepfake* es discurso, y, como tal, es y deben ser igualmente digno de debate los polimorfismos del *deepfake* de la libertad de expresión (Lendvai, 2023) y la desinformación. Y estas polémicas conllevan cuestiones sobre posibles respuestas legales y soluciones reguladoras y mejores prácticas.

Siguiendo con la conceptualización, y para comprender mejor las filosofías jurídicas que subyacen a la normativa, es importante destacar los casos y las polémicas clave en torno a la relación entre desinformación y *deepfake*. En el siguiente segmento, subrayamos el trasfondo teórico de la desinformación *deepfake*, con especial atención a la desinformación política.

2. Deepfake y desinformación

2.1. ¿*Hack* transparente o arma pesada de desinformación?

En marzo de 2022 se produce un ataque bien conocido y contemporáneo (Espaliú-Bedud, 2023), a saber, uno basado en el uso del *deepfake* por parte de Rusia. El presidente ucraniano Volodymyr Zelensky pronuncia un discurso en un vídeo que se difunde como la pólvora en las redes sociales ucranianas y rusas: en la confusa grabación, Zelensky pone una cara extraña y pide a los soldados ucranianos que depongan las armas (Allyn, 2022). Contenido accesible, visible, que parece real, pero es falso.

Según Rob Cover (2022, pág. 615), el *deepfake* es ante todo una «preocupación social». Y esta preocupación social es principalmente la tesis del artículo de Adam Satarino y Paul Mozur (2023): la gente es falsa, pero la desinformación es real. El *deepfake* tiene efectos controvertidos en diversos ámbitos, tanto jurídicos como de la esfera pública democrática; provoca controversias sobre el uso de la identidad y la imagen y tiene el potencial de influir negativamente en la opinión pública (Van der Sloot y Wagensveld, 2022). En este contexto, también hay que destacar el uso de *deepfakes* para el engaño político o incluso en conflictos armados, como se menciona en el párrafo anterior (Allen, 2022). Andrew Ray subraya (2021) que el contenido *deepfake* puede contribuir en gran medida a una disminución de la confianza política, y que también puede suponer una amenaza real de interferencia ilegal en las elecciones.

En palabras de Ágnes Veszelszki (2022), el contenido *deepfake* es una «verdadera arma pesada». Aunque las definiciones difieren (Fathaigh *et al*., 2021), desinformación, especialmente en el contexto del *deepfake*, significa la difusión deliberada de información falsa o engañosa con el objetivo de manipular estratégicamente a una audiencia o crear más fracturas en las divisiones existentes (Fallis, 2015; Arce, 2024). Este fenómeno se ve exacerbado por la sofisticación tecnológica de los *deepfakes* que hace que el contenido falso sea más convincente (Levak, 2021). Debido a esto, uno de los mayores retos en la lucha contra las *fake news*, además de la creación de noticias falsas en sí, es que son generadas por *deepfake* y su difusión depende no tanto de usuarios con intenciones desinformativas como sí lo hace mayoritariamente de usuarios que se dejan engañar por el contenido manipulado. La investigación de Vaccari y Chadwick (2020) evaluó la capacidad de engaño de los *deepfakes*. Un vídeo de 4 segundos engañó al 15 % de los participantes, y un 35 % no estaba seguro de su autenticidad. Un vídeo de 26 segundos aumentó estas cifras al 16 % engañados y al 37 % inseguros. La incertidumbre generada por los *deepfakes* es alarmante: si la tecnología de hace tres años causaba confusión en más de la mitad de las personas, es razonable suponer que la tecnología de 2024 podría generar aún más engaño e incertidumbre en los consumidores.

En el presente estudiamos cómo puede regular la legislación esta preocupación social prevalente. En la siguiente sección, investigamos los distintos tipos de enfoques reguladores, a saber, el enfoque liberal (Estados Unidos) y el enfoque social/basado en el riesgo (Unión Europea).

2.2. DEEPFAKES Act y AI Act: respuestas legislativas a los *deepfake* content y las *deepfake fake news*

Andrew Ray (2021, págs. 991-992) esboza tres problemas normativos que plantea la regulación de los *deepfakes* políticos engañosos. El primer problema, según el autor, es que la eliminación del contenido *deepfake* no resuelve los efectos perjudiciales del contenido del mismo: la eliminación no tiene como resultado la mejora de la posición jurídica de la víctima respecto del efecto perjudicial ni la restauración de su reputación. El segundo problema es que no todas las plataformas de medios sociales tratan el *deepfake* como contenido de desinformación. El tercer problema se refiere a la conceptualización del *deepfake*: la definición imprecisa de este hace que la legislación no pueda proteger eficazmente a los afectados. Siguiendo estas tres líneas de preocupación, la siguiente sección esboza las tendencias normativas que pueden observarse, especialmente en Estados Unidos y la Unión Europea.

2.2.1. Soluciones americanas

En el contexto de la normativa estadounidense sobre *deepfake*, es importante analizar brevemente las prácticas diferentes y a menudo divergentes de los Estados miembros. El enfoque estadounidense es muy liberal. Un ejemplo frecuente de ello es el hecho de que, desde 1996, la Sección 230 de la Ley de Decencia en las Comunicaciones (CDA) garantiza que los intermediarios están exentos de responsabilidad por los contenidos, incluso los ilícitos, compartidos a través de ellos por terceros. Aunque en la actualidad no existe en Estados Unidos una normativa federal sobre el tratamiento perjudicial de los contenidos *deepfake*, varios Estados miembros cuentan con normativas particulares y, en algunos casos, muy específicas (solo relacionadas con las elecciones) sobre *deepfake*. En el contexto de la CDA, esto significa que ciertas medidas legislativas intentan revisar las «Veintiséis Palabras que Crearon Internet» –una referencia a la estipulación de las 26 palabras de la Sección 230 de la CDA: «Ningún proveedor o usuario de un servicio informático interactivo será tratado como editor o altavoz de cualquier información facilitada por otro proveedor de contenidos informativos.»

En 2019, la representante Yvette Clarke presentó al Congreso el primer proyecto de ley sobre *deepfake* de ámbito superior a las regulaciones estatales, la DEEPFAKES Accountability Act (H. R. 3230 § 1041), que habría proporcionado un medio para que las partes afectadas por

contenidos *deepfake* dañinos pudieran reclamar daños y propuso una definición uniforme de deepfake (H.R. 3230 § 1041(n)(3)): «Una *deepfake* es cualquier grabación de vídeo, película, grabación de sonido, imagen electrónica o fotografía, o cualquier representación tecnológica de discurso o conducta, que A) parezca ser una representación fiel del discurso o conducta de una persona que no ha participado realmente en dicho discurso o conducta; y B) sea de hecho producida por medios técnicos y no por la capacidad de otra persona de hacerse pasar por dicha persona física u oralmente.»

Sin embargo, el uso del tiempo pasado condicional no es casual; la propuesta de Clarke fracasó en el Congreso. Aunque la necesidad de regular el *deepfake* a nivel federal estadounidense no ha desaparecido: unas semanas antes de escribir este artículo, Joe Biden firmó la Orden Ejecutiva sobre Inteligencia Artificial en octubre de 2023, que pretende ser la primera del mundo en abordar los problemas causados por la IA, incluido el *deepfake* (Johnson, 2023). Aunque hasta ahora no se ha aplicado ningún instrumento jurídico tras la orden ejecutiva, es importante subrayar que esta pretende hacer frente a los retos del deepfake mediante una tecnología avanzada de marcas de agua, una estipulación que ya se ha enfrentado a duras críticas en cuanto a su eficacia (Strickland, 2023).

Entre las normativas de los estados de Estados Unidos, destaca el caso de California, donde entraron en vigor dos reglamentos específicos sobre *deepfake* en 2020: uno para la falsificación de material de campaña utilizando tecnología *deepfake*, mientras que el otro establece un derecho de acción privado contra el productor de contenido creado artificialmente que, a sabiendas e intencionadamente, cree contenido sexual sobre otra persona sin su consentimiento. En el contexto de los materiales de campaña, se promulgó el Código Electoral de California §20010, que permite a un candidato a un cargo político interponer una acción dentro de los 60 días siguientes a las elecciones contra una persona que haya producido contenido *deepfake* con la intención de «inculpar» al candidato durante la campaña. Siguiendo las ideas de Rob Cover (2022) a este respecto, cabe señalar que la enmienda californiana aborda una cuestión crítica: como se ha expuesto repetidamente en la subsección 2.1, el contenido *deepfake* malicioso tiene el potencial de causar un gran daño a la carrera de una persona con aspiraciones políticas. Sin embargo, es cuestionable si el reglamento, suponiendo que sea aplicable, es decir, que se encuentre al productor del material difamatorio, remediará el daño al candidato afectado por la campaña de difamación.

Volviendo al primer problema de Andrew Ray, la legislación californiana ofrece una solución parcial para el afectado: si bien el candidato puede emprender acciones contra el proveedor de contenidos, la disposición no proporciona ningún alivio real en caso de daño a su reputación o incluso de una posible derrota electoral.

El vínculo entre la regulación de la desinformación y el *deepfake* es más distante y general en Virginia, donde la Unlawful Dissemination or Sale of Images of Another Person Act (Ferraro 2019: 14; Va. Code Ann. § 18.2-386.2; HB 2678, SB 1736 2019) tipifica como delito menor la difusión de imágenes y vídeos de otra persona generados sin consentimiento (es decir, *deepfakes*) (HB 2678 VA 2019), con penas para este tipo de delitos que van desde una multa hasta un año de prisión.

2.2.2 La regulación de la inteligencia artificial y el enfoque deepfake de la UE: bajo riesgo, pocas limitaciones

En la regulación europea, la propuesta del reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA (Ley de Inteligencia Artificial, en lo sucesivo LIA) debería mencionarse como prioritaria. El reglamento propuesto no ofrece una definición precisa de *deepfake*, pero sí lo regula como un sistema de IA de bajo riesgo. En consecuencia, los contenidos *deepfake* están cubiertos por el artículo 52, que impone obligaciones generales de transparencia a los proveedores de servicios. Según el artículo 52, apartado 3, los usuarios de sistemas de IA que generen o manipulen imágenes, contenidos de audio o vídeo que tengan un parecido perceptible con personas, objetos, lugares u otras entidades o acontecimientos existentes que puedan parecer a una persona engañosamente genuinos o reales («*deepfake*»), deben revelar que el contenido ha sido creado o manipulado artificialmente. Aunque el texto definitivo no se ha publicado en el Diario Oficial de la UE, la última versión que circula sobre la Ley de IA destaca específicamente el *deepfake* en un nuevo contexto (Parlamento Europeo, 2024). Según la última versión no consolidada, el artículo 50 (2) estipula que los proveedores de sistemas de IA, incluidos los sistemas de propósito general que generan contenidos sintéticos de audio, imagen, vídeo o texto, deben garantizar que los resultados

estén marcados en un formato legible por máquina para indicar su origen artificial. Las soluciones técnicas deben ser eficaces, interoperables, sólidas y fiables, teniendo en cuenta las particularidades de cada tipo de contenido, los costes de aplicación y el estado actual de la técnica.

Esta postura reguladora, es decir, la inclusión en la legislación de *deepfake* como una tecnología de IA de bajo riesgo, es muy sorprendente y aún más controvertida, ya que no hay una explicación lógica de por qué, a la luz del hecho de que se ha descubierto tanto en círculos académicos como profesionales que el contenido *deepfake* es casi en un 90 % contenido pornográfico no consentido (Hao, 2021), los desarrolladores no están sujetos a las normas de transparencia más mínimas.

Aunque no se aplica directamente a la tecnología *deepfake*, es importante destacar el reglamento 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 sobre el mercado único de servicios digitales, que modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales o DSA). El DSA busca promover la soberanía digital europea, enfocándose en derechos fundamentales, privacidad de datos y protección de las partes interesadas (Turillazzi *et al*., 2023), regulando la evaluación de riesgos por grandes plataformas en línea como Facebook, Instagram o TikTok (artículos 34-36, incluidos mecanismos de respuesta a crisis). El DSA impone obligaciones asimétricas para grandes plataformas. Según el artículo 34, estas deben identificar, analizar y evaluar los riesgos sistémicos de sus servicios, incluidos los sistemas algorítmicos, antes de una fecha específica y anualmente, especialmente antes de desplegar funciones que puedan afectar significativamente los riesgos identificados. Estos riesgos incluyen la difusión de contenidos ilegales, efectos negativos sobre derechos fundamentales (como la dignidad humana, privacidad, protección de datos, libertad de expresión y no discriminación), impactos en el discurso cívico y la seguridad pública, y cuestiones relacionadas con violencia de género, salud pública y bienestar. El considerando 80 del DSA identifica cuatro categorías de riesgos sistémicos.

1) La primera se refiere a la difusión de contenidos ilegales, como la incitación al odio y materiales de abuso sexual infantil.

2) La segunda categoría se refiere al impacto de grandes plataformas en la libertad de los medios, protección de datos y derechos fundamentales, con riesgos de sesgos algorítmicos e interfaces perjudiciales.

3) La tercera categoría aborda los riesgos que socavan procesos democráticos y discurso público.

4) La cuarta se refiere a la desinformación sobre salud pública y bienestar de los usuarios, relevante en la difusión de noticias falsas en redes sociales.

La relevancia del DSA en relación con la desinformación mediante *deepfakes* radica en que todas las categorías de riesgo sistémico pueden originarse en campañas de desinformación usando *deepfakes*: un vídeo falso de un político llamando a la guerra, un científico anunciando una nueva pandemia o contenido propagandístico para incitar al odio contra minorías. El DSA es una piedra angular de la regulación horizontal, obligando a grandes plataformas a actuar, mitigar y evaluar riesgos para salvaguardar valores democráticos y usuarios (Gosztonyi, 2023).

A pesar de que parece existir una normativa aplicable a la desinformación mediante *deepfake*, la tendencia a difundir desinformación a través de esta tecnología no se ha frenado (Gambín *et al*., 2024). En la siguiente sección se analizan resoluciones «fuera de la legislación» y se citan documentos que, aunque no tienen carácter vinculante, pueden ser de gran utilidad para futuras normativas.

2.2.3. ¿Más allá del occidentalismo? –breve resumen de la perspectiva china de la «regulación control»

Para alejarnos de los enfoques occidentalistas del *deepfake* es importante incluir brevemente un aspecto novedoso sobre la IA y el *deepfake*, a saber, el enfoque chino de «control» en la regulación. En el contexto de China, cabe destacar que existen normativas alternativas más estrictas sobre *deepfake*. El 28 de enero de 2022, la Administración del Ciberespacio de China publicó una serie de normas que regulan la administración de los servicios de información por internet que utilizan la síntesis *deepfake*. La normativa nacional se titula Ley de Servicio de Información por Internet Reglamento de Tecnología de Síntesis Profunda y fue promulgada el 25 de noviembre de 2022, lo que la convierte en el primer instrumento jurídico nacional dirigido específicamente a los contenidos *deepfake*. El término «tecnología de síntesis profunda» en este reglamento (§2) abarca ampliamente casi todos los contenidos audiovisuales. Entre las principales estipulaciones figuran las dirigidas a los proveedores de servicios de *deepfake* con requisitos de ciberseguridad, verificación del nombre real, gestión de datos y etiquetado obligatorio de los con-

tenidos sintéticos –estas medidas están diseñadas para frenar la creación y difusión de información engañosa. La ley pretende impedir la difusión de contenidos falsos convincentes que pueden socavar la estabilidad social al propagar desinformación (Hine y Floridi, 2022). Aunque este nuevo enfoque legal ha recibido una acogida positiva, ya que se espera que ofrezca una mayor protección a los usuarios de internet y aplique una responsabilidad legal más estricta a las empresas que desarrollan tecnología de *deepfake*, Hine y Floiridi (2022) destacan las dificultades técnicas de su aplicación: como garantizar la permanencia de las etiquetas en los contenidos sintéticos, ya que las marcas de agua pueden eliminarse y los metadatos alterarse. Además, controlar la difusión de contenidos de síntesis profunda es intrínsecamente difícil; una vez creados, estos contenidos pueden desvincularse fácilmente de su origen y difundirse de forma independiente, lo que complica los esfuerzos por rastrearlos y eliminarlos por completo de internet.

Para una mejor comprensión de los tres enfoques legislativos, ofrecemos el siguiente cuadro:

Tabla 1. Comparativa de los enfoques legislativos

	Estados Unidos	*Unión Europea*	*China*
Situación normativa actual y filosofía reguladora	Enfoque normativo muy liberal, normativa fragmentada y ausencia de normativa federal vinculante.	Gran atención a la transparencia, directrices claras sobre temas como la desinformación, normativa vinculante de próxima aparición (LIA), multitud de instrumentos no vinculantes (Código sobre Desinformación). Alto nivel de multilateralismo, tendencia a la regulación horizontal.	Regulación más bien estricta, atención específica a los contenidos *deepfake* nocivos, centrada en el desarrollo de tecnologías sintéticas.
Transparencia	Bajos niveles de transparencia, fuerte tendencia a impulsar la innovación.	Niveles adecuados de transparencia que pueden seguir las empresas. Plazos adecuados para informar, sin embargo, las medidas de transparencia exhaustivas son obligatorias.	Los proveedores de servicios de síntesis profunda están obligados a establecer directrices, criterios y procedimientos para identificar la información falsa o perjudicial y gestionar a los usuarios que creen tales contenidos con la tecnología de síntesis profunda. (Interesse, 2022)
Objetivos	Una regulación que sea lo suficientemente liberal como para permitir a las empresas desarrollar tecnologías de IA sin problemas sustanciales de conformidad jurídica.	Una regulación exhaustiva que pueda servir de «modelo» mundial para que otros países la sigan. Medidas rígidas para garantizar una experiencia centrada en el usuario con la IA.	Cuatro objetivos: 1) Seguridad de los datos y protección de la información personal, 2) Transparencia, 3) Gestión de contenidos, y 4) etiquetado y seguridad técnica (Interesse, 2022).
Apreciaciones específicas sobre *deepfake* desinformación y libertad de expresión	Enfoque liberal que regula los contenidos deepfake (como forma de expresión) con normas muy liberales	Enfoque moderado que regula el *deepfake* content (como forma de expresión) mediante medidas de transparencia y diligencia debida por parte de las empresas. En la práctica, es probable que el Tribunal Europeo de Derechos Humanos sea importante a la hora de trazar el camino exacto.	Enfoque estricto; por ejemplo, el establecimiento de un mecanismo para contrarrestar las noticias falsas ordena que cuando se utilicen servicios de síntesis profunda para crear, replicar, publicar o difundir información falsa, los proveedores de servicios deben tomar medidas para disipar esas noticias, mantener registros e informar de los incidentes a las autoridades pertinentes.
Filosofía reguladora	Recurre a medidas *ex post*, abordando los problemas *a posteriori* mediante litigios y leyes estatales.	Con énfasis en la evaluación de riesgos y la transparencia, se inclina hacia la regulación *ex ante*, con el objetivo de prevenir los daños antes de que se produzcan. Las medidas *ex post* incluyen sanciones.	Enfoque centrado, basado en el control y en tecnologías específicas.

Fuente: elaboración propia.

3. Alternativas y perspectivas al margen de la regulación

La realidad del problema se ilustra mejor en las declaraciones emitidas en julio de 2019 por el Relator Especial de la ONU para la Libertad de Expresión, el Relator Especial de la Organización para la Seguridad y la Cooperación en Europa (OSCE) para la Libertad de Prensa, el Relator Especial de la Organización de Estados Americanos (OEA) para la Libertad de Expresión y el Relator Especial de la Comisión Africana de Derechos Humanos y de los Pueblos (CADHP) para la Libertad de Expresión y el Acceso a la Información. Todos ellos expusieron los retos a los que se enfrenta la libertad de expresión en la próxima década y pidieron a los principales políticos del mundo y a los propietarios de plataformas en línea que adopten soluciones que tengan en cuenta los derechos humanos. Destacaron los retos que plantea la desinformación, incluida la creciente aparición del *deepfake* (Joint Declaration of the UN, OSCE, OAS and ACHPR, 2019, pág. 3).

Ofrecemos tres soluciones a sendos problemas de Andrew Ray mencionados anteriormente. Por lo que respecta a la eliminación de contenidos *deepfake*, ni la solución estadounidense ni la europea ofrecen una verdadera asistencia jurídica. En este contexto, se consideran posibles dos soluciones. La solución *ex ante* podría consistir en obligar a las plataformas a utilizar sistemas de detección de *deepfakes* que vayan evolucionando y sigan la evolución dinámica de la democratización de los mismos (Masood *et al*., 2022), es decir, que las plataformas añadan automáticamente una marca en los vídeos después de publicar el post que indique que el vídeo se ha realizado con tecnología *deepfake*. Sin embargo, en lo que respecta a la solución ex post, contrariamente a la polémica de Ray, creemos que las herramientas legales no están bien adaptadas para intervenir eficazmente más allá de la indemnización ex post de daños y perjuicios para restaurar la reputación de un candidato o una persona implicada en la desinformación *deepfake* –un punto que se asemeja mucho a las críticas generales de los procedimientos de reivindicación de la prensa.

Sugerimos, sin embargo, que la solución a este problema pasa por la educación mediática y el fomento del pensamiento crítico: comprobar a posteriori que lo que dice un político suena «poco creíble» puede ser una forma más obvia de resolver la polémica que confiar en la lenta molienda de los molinos de la regulación. También nos gustaría destacar el posible papel de los fact-checkers, que podrían comprobar la autenticidad de los contenidos *deepfake* a corto plazo, incluso utilizando plataformas. En cuanto a la segunda cuestión planteada por Ray, es decir, la armonización de los conceptos y fenómenos de desinformación entre plataformas, podemos mencionar el Código de Buenas Prácticas sobre Desinformación 2022 de la Unión Europea, cuyo objetivo es proporcionar un marco común para la cooperación entre plataformas en materia de actividades de desinformación y buenas prácticas (Espaliú-Berdud, 2024). El Código, además de identificar opciones conceptuales y prácticas comunes, es una de las iniciativas en las que participan activamente casi todas las plataformas en línea.

También es importante subrayar que, especialmente las Naciones Unidas (2023) y el G7 (2023), están desarrollando activamente el discurso sobre la disinformación, la IA y la regulación de los *deepfakes* a través de informes y resoluciones. Estos informes contribuyen activamente a desarrollar una regulación y aplicación mejore y más transparente de la IA, proponiendo un enfoque basado en el ser humano y un «multistakeholderismo» de alto nivel que es más que deseable, ya que fomenta aún más la democratización de la legislación y la rendición de cuentas por parte de las empresas (Prem, 2020).

Conclusiones

Por todo lo anterior, parece claro que el derecho como «regulador único y particularista» no es suficiente para luchar contra los contenidos ilegales en línea. El espectacular y eficaz desarrollo de los sistemas de detección de *deepfakes* es un avance bienvenido, así como la serie de disposiciones supranacionales (la LIA o el DSA) que van más allá de las normativas nacionales –a menudo divergentes–, pero también creemos que la concienciación social es esencial. Se propone poner en marcha un proceso de concienciación y formación de los medios de comunicación para educar a los internautas no solo sobre los peligros del *deepfake*, sino también sobre las medidas que pueden tomar para identificar este tipo de contenidos. Por lo tanto, esto requiere un enfoque holístico en el que los actores del «triángulo de regulación de plataformas» (Gorwa, 2019) tomen medidas conjuntamente para abordar los problemas que surgen, no solo a través de las herramientas de la legislación, sino también a través del desarrollo de la alfabetización mediática general.

Referencias bibliográficas

2018 CODE OF PRACTICE ON DISINFORMATION (2022). *Comisión Europea* [en línea]. Disponible en: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation

ALLEN, M. D. N. (2022). «Deepfake Fight: AI-Powered Disinformation and Perfidy Under the Geneva Conventions». *Journal of Emerging Technologies*, vol. 3, n.º 2, págs. 1-70. DOI: https://doi.org/10.2139/ssrn.3958426

ALLYN, B. (2022). «Deepfake Video of Zelenskyy Could Be 'Tip of the Iceberg' in Info War, Experts Warn». *NPR*, [en línea]. Disponible en: https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia/. [Fecha de consulta: 6 de marzo de 2024].

ARCE, D. (2024) «Disinformation Strategies». *Defence and Peace Economics*, págs. 1-14. DOI: https://doi.org/10.1080/10242694.2024.2302236

CALIFORNIAN ELECTIONS CODE §20010 [en línea] Disponible en: https://ocvote.gov/apps/legtracker/elections-code/contents/.

CITRON, D. K.; CHESNEY, R. (2019). «Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security». *California Law Review*, n.º 107, págs. 1753-1820. DOI: https://doi.org/10.2139/ssrn.3213954

CÓDIGO ELECTORAL DE CALIFORNIA §20010. [en línea]. Disponible en: https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=20010.&nodeTreePath=22.1&lawCode=ELEC

COE, P. (2023). «Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory». *Journal of Media Law*, vol. 15, n.º 2, págs. 213-242. DOI: https://doi.org/10.1080/17577632.2024.2316360

COLE, S. (2017). «AI-Assisted Fake Porn Is Here and We're All Fucked». *Vice*, [en línea]. Disponible en: https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn/. [Fecha de consulta: 6 de marzo de 2024].

Communications Decency Act of 1996 (CDA), 47 U.S.C. § 230.

COVER, R. (2022). «Deepfake Culture: The Emergence of Audio-video Deception as an Object of Social Anxiety and Regulation». *Continuum Journal of Media & Cultural Studies*, n.º 36, págs. 609-621. DOI: https://doi.org/10.1080/10304312.2022.2084039

H.R.5586 - DEEPFAKES Accountability Act, 18th Congress (2023-2024), Introduced in House (20.09.2023)

DICKINSON, D. (2018). «Interview with Farnaz Fassihi». *UN News*, [en línea]. Disponible en: https://news.un.org/en/audio/2018/05/1008682/. [Fecha de consulta: 6 de marzo de 2024].

ESPALIÚ-BERDUD, C. (2022). «Legal and criminal prosecution of disinformation in Spain in the context of the European Union». *Profesional De La información The Information Professional*, vol. 31, n.º 3, págs. 1-14. DOI: https://doi.org/10.3145/epi.2022.may.22

ESPALIÚ-BERDUD, C. (2023). «Use of disinformation as a weapon in contemporary international relations: accountability for Russian actions against states and international organizations». *Profesional De La información The Information Professional*, vol. 32, no. 4, págs. 1-19. →DOI: https://doi.org/10.3145/epi.2023.jul.02

ESPALIÚ-BERDUD, C. (2024). «The EU Code of Practice on Disinformation: An Example of the Self-Regulatory Trend in International and European Law». *VISUAL REVIEW – International Visual Cul-*

ture Review Revista Internacional De Cultura Visual, vol. 16, n.º 2, págs. 95-109. DOI: https://doi.org/10.62161/revvisual.v16.5217

FALLIS, D. (2015). «What Is Disinformation?». *Library Trends*, vol. 63, n.º 3, págs. 401-426. DOI: https://doi.org/10.1353/lib.2015.0014

FATHAIGH, R.Ó.; HELBERGER, N.; APPELMAN, N. (2021). «The Perils of Legally Defining Disinformation». *Internet Policy Review*, vol. 10, n.º 4, págs. 1-25. DOI: https://doi.org/10.14763/2021.4.1584

FERRARO, M. F. (2019). «Deepfake Legislation: A Nationwide Survey – State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media». *WilmerHale report*, 2019.

GAMBÍN, N. F.; YAZIDI, A.; VASILAKOS, A.; HAUGERUD, H.; DJENOURI, Y. (2024). «Deepfakes: current and future trends». *Artificial Intelligence Review*, vol. 57, n.º 3. DOI: https://doi.org/10.1007/s10462-023-10679-x

GORWA, R. (2019). «The Platform Governance Triangle: Conceptualising the Informal Regulation of Online Content». *Internet Policy Review*, n.º 8, págs. 1-20. DOI: https://doi.org/10.14763/2019.2.1407

GOSZTONYI, G. (2023). *Censorship from Plato to Social Media. The Complexity of Social Media's Content Regulation and Moderation Practices*. Cham: Springer Nature Switzerland AG. DOI: https://doi.org/10.1007/978-3-031-46529-1

G7 (2023). «Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system». *Comisión Europea* [en línea]. Disponible en: https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-guiding-principles-advanced-ai-system

HAO, K. (2021). «Deepfake Porn is Ruining Women's Lives. Now the Law May Finally Ban it». *MIT Technology Review* [en línea]. Disponible en: https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/. [Fecha de consulta: 6 de marzo de 2024].

HINE, E.; FLORIDI, L. (2022). «New deepfake regulations in China are a tool for social stability, but at what cost?». *Nature Machine Intelligence*, n.º 4, págs. 608-610. DOI: https://doi.org/10.1038/s42256-022-00513-4

HUSA, J. (2022). «Interdisciplinary comparative law: Rubbing shoulders with the neighbours or standing alone in a crowd». *Internet Policy Review*, n.º 8, págs. 1_20. DOI: https://doi.org/10.14763/2019.2.1407

HB 2678, SB 1736 2019, *LIS* [en línea]. Disponible en: https://lis.virginia.gov/cgi-bin/legp604.exe?191+sum+HB2678

HB 2678 VA 2019, *LIS* [en línea]. Disponible en: https://lis.virginia.gov/cgi-bin/legp604.exe?191+sum+SB1736 DOI: https://doi.org/10.1007/s40278-019-57084-4

H.R. 3230 § 1041, *Congress.gov* [en línea]. Disponible en: https://www.congress.gov/bill/116th-congress/house-bill/3230

H.R. 3230 § 1041(n)(3), *Congress.gov* [en línea]. Disponible en: https://www.congress.gov/bill/116th-congress/house-bill/3230

INTERESSE, G. (2022). «China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023». *China Briefing*, [en línea]. Disponible en: https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/. [Fecha de consulta: 25 de mayo de 2024].

IPSOS (2022). «Internet Users' Trust in the Internet Has Dropped Significantly Since 2019». *Ipsos*, [en línea]. Disponible en: https://www.ipsos.com/en/trust-in-the-internet-2022/. [Fecha de consulta: 6 de marzo de 2024].

JOHNSON, T. (2023). «Joe Biden Talks About Watching An AI Generated Deepfake Of Himself: "I Said, When The Hell Did I Say That?"». *Deadline*, [en línea]. Disponible en: https://deadline.com/2023/10/ai-joe-biden-executive-order-1235586979/. [Fecha de consulta: 6 de marzo de 2024].

Joint Declaration of the UN, OSCE, OAS and ACHPR (2019), [en línea]. Disponible en: https://www.osce.org/files/f/documents/9/c/425282.pdf

LENDVAI, G. F. (2023). «Deepfake a szólásszabadság tükrében: Reflexiók a jog perspektívájából». En: *Deepfake: a valótlan valóság*, págs. 121-138. Budapest: Gondolat Kiadó,

LEVAK, T. (2021). «Disinformation in the new media system – Characteristics, forms, reasons for its dissemination and potential means of tackling the issue». *Medijska Istraživanja*, vol. 26, n.º 2, págs. 29-58. DOI: https://doi.org/10.22572/mi.26.2.2

Ley de Servicio de Información por Internet Reglamento de Tecnología de Síntesis Profunda (国家互联网信息办公室等三部门发布《互联网信息服务深度合成管理规定, 11.12.2022, *CAC* [en línea]. Disponible en: https://www.cac.gov.cn/2022-12/11/c_1672221949318230.htm

MASOOD, M. *et al*. (2022). «Deepfakes Generation and Detection: State-of-the-art, Open Challenges, Countermeasures, and Way Forward». *Applied Intelligence*, n.º 53, págs. 1-53. DOI: https://doi.org/10.1007/s10489-022-03766-z

NACIONES UNIDAS (2023). *Governing AI for Humanity*. Informe de Diciembre de 2023 [en línea]. Disponible en: https://www.un.org/sites/un2.un.org/files/ai_advisory_body_interim_report.pdf

Online Safety Act 2023 (Ley de Seguridad en Línea de 2023), UK Public General Acts, 2023.c.50., 26.10.2023.

PANTSEREV, K. A. (2020). «The Malicious Use of AI-Based Deepfake Technology as the New Threat to Psychological Security and Political Stability». En: *Cyber Defence in the Age of AI, Smart Societies and Augmented Humanity*, págs. 47-48. Cham: Springer Nature Switzerland AG. DOI: https://doi.org/10.1007/978-3-030-35746-7_3

PARLAMENTO EUROPEO (2024). «Corrección de errores del Reglamento de Inteligencia Artificial» [en línea]. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_ES.pdf. [Fecha de consulta: 24 de mayo de 2024].

PREM, B. (2020). «The False Promise of Multi-stakeholder Governance: Depoliticising Private Military and Security Companies». *Global Society*, n.º 35, págs. 149-170. DOI: https://doi.org/10.1080/13600826.2020.1791055

RANKIN, J. (2018). «Viktor Orbán: Re-election of Hungary's Anti-immigrant Leader is Major Challenge for EU». *The Guardian*, [en línea]. Disponible en: https://www.theguardian.com/world/2018/apr/09/viktor-orban-re-election-hungarys-anti-immigrant-leader-major-challenge-for-eu/. [Fecha de consulta: 6 de marzo de 2024].

RAY, A. (2021). «Disinformation, Deepfakes and Democracies: The Need for Legislative Reform». *UNSW Law Journal*, n.º 44, págs. 983-1013. DOI: https://doi.org/10.53637/DELS2700

Reglamento (UE) 2022/2065 del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales) (El DSA), PE/30/2022/REV/1, OJ L 277, 27/10/2022, págs. 1-102.

Proquesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican determinados actos legislativos de la unión (Ley de Inteligencia Artificial, en lo sucesivo: LIA), COM/2021/206 final (21.4.2021)

SATARINO, A.; MOZUR, P. (2023). «The People on Screen are Fake. The Disinformation is Real». *The New York Times*, [en línea]. Disponible en: https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html. [Fecha de consulta: 6 de marzo de 2024].

SPINA, E.L. (2024). «Between the Need and the Purpose of Comparing in Legal Research. Reflections on its Methodological Approach». *Anales de la Cátedra Francisco Suarez*, n.º 58, págs. 47-71. DOI: https://doi.org/10.30827/acfs.v58i.28723

STRICKLAND, E. (2023). «What You Need to Know About Biden's Sweeping AI Order». *IEEE Spectrum*, [en línea]. Disponible en: https://spectrum.ieee.org/biden-ai-executive-order. [Fecha de consulta: 24 de mayo de 2024].

TURILLAZZI, A.; TADDEO, M.; FLORIDI, L.; CASOLARI, F. (2023). «The Digital Services Act: An Analysis of Its Ethical, Legal, and Social Implications». *Law, Innovation and Technology*, vol. 15, n.º 1, págs. 83-106. DOI: https://doi.org/10.1080/17579961.2023.2184136

Unlawful Dissemination or Sale of Images of Another Person Act, Code of Virginia, 2019, cc. 490, 515.

VACCARI, C.; ANDREW, C. (2020). «Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News». *Social Media + Society*, n.º 6, págs. 1-13. DOI: https://doi.org/10.1177/2056305120903408

VAN DER SLOOT, B.; WAGENSVELD, Y. (2022). «Deepfakes: Regulatory Challenges for the Synthetic Society». *Computer Law & Security Review*, n.º 46, págs. 1-15. DOI: https://doi.org/10.1016/j.clsr.2022.105716

VESZELSZKI, Á. (2022). «A tudományos influencerektől a deepfake-ig. A legújabb tudománykommunikációs lehetőségek». *Filológia*, n.º 13, págs. 27-39.

Cita recomendada

LENDVAI, GERGELY FERENC; GOSZTONYI, GERGELY (2024). «*Deepfake* y desinformación – ¿Qué puede hacer el derecho frente a las noticias falsas creadas por *deepfake*?». *IDP. Revista de Internet, Derecho y Política*, núm. 41. UOC. [Fecha de consulta: dd/mm/aa]. DOI: http://dx.doi.org/10.7238/idp.v0i41.427515

Sobre las autorías

Gergely Ferenc Lendvai
Universidad Católica Pázmány Péter
ergelyflendvai@gmail.com

Doctorando en la Universidad Católica Pázmány Péter e investigador en el Centro de Derecho de la Sociedad de la Información de la Universidad de Milán y la Facultad de Artes y Ciencias de la Universidad de Richmond. La investigación de Gergely se centra en los aspectos transdisciplinares del Derecho, con especial atención a los aspectos sociológicos, la marginación en línea, la incitación al odio contra las minorías, las zonas grises tecnológicas y las investigaciones jurídicas empíricas. Gergely trabaja bajo una beca de la Fundación Rosztoczy.

Gergely Gosztonyi
Universidad Eötvös Loránd (ELTE)
gosztonyi@ajk.elte.hu

Profesor asociado habilitado, abogado húngaro e investigador de medios de comunicación. Se licenció en la Facultad de Derecho y Ciencias Políticas de la Universidad Eötvös Loránd (ELTE), y desde entonces imparte clases en la misma institución: imparte diversos cursos sobre derecho de los medios de comunicación, derecho constitucional e historia jurídica a nivel de licenciatura, máster y doctorado. Sus intereses de investigación incluyen la regulación global de los medios sociales, la censura, el deepfake, los medios alternativos y la responsabilidad de los intermediarios. Desde 2015, ha sido el entrenador principal del equipo húngaro para el Monroe E. Price Media Law Moot Court Competition. Su libro más reciente es *Censorship from Plato to Social Media. The Complexity of Social Media's Content Regulation and Moderation Practices* (Springer Nature Switzerland AG, 2023).