



Revista Galega de Administración Pública, EGAP
Núm. 68_julio-diciembre 2024 | pp. 311-336
Santiago de Compostela, 2024
<https://doi.org/10.36402/regap.v68i1.5240>
© Luisa Irene González Martínez
ISSN-e: 1132-8371 | ISSN: 1132-8371
Recibido: 15/11/2024 | Aceptado: 27/03/2025

Editado bajo licencia Creative Commons Attribution 4.0 International License

Vulnerabilidade dos procesos electorais ante o mal uso da intelixencia artificial

Vulnerabilidad de los procesos electorales ante el mal uso de la inteligencia artificial

The vulnerability of electoral processes in the face of AI misuse

68 Regap

Regap



ESTUDIOS

LUISA IRENE GONZÁLEZ MARTÍNEZ
Investigadora predoctoral de Derecho Constitucional
Universidad de A Coruña
i.gonzalez.martinez@udc.es

ORCID: 0009-0000-6704-2684.

Resumo: Os procesos electorais democráticos susténtanse na libre formación da opinión pública, baseada en información veraz, plural e contrastada, e expresada libremente mediante o voto. Non obstante, nun ámbito onde resulta cada vez máis difícil discernir entre información veraz e desinformación, e onde as redes sociais están a substituír os medios de comunicación tradicionais como fontes informativas, xorde a interrogante de se a formación da opinión pública continúa sendo verdadeiramente libre. Esta preocupación intensifícase co avance da intelixencia artificial, que permite esvaecer a verdade e condicionar a liberdade de elección mediante *deepfakes*, propaganda computacional e segmentación algorítmica, presentando un novo desafío para o cal o noso sistema electoral podería non estar axeitadamente preparado.

Palabras clave: Liberdade de expresión, liberdade de información, opinión pública, eleccións, intelixencia artificial, desinformación, *deepfake*, Cambridge Analytica.

Resumen: Los procesos electorales democráticos se sustentan en la libre formación de la opinión pública, basada en información veraz, plural y contrastada, y expresada libremente mediante el voto. Sin embargo, en un entorno donde resulta cada vez más difícil discernir entre información veraz y desinformación, y donde las redes sociales están reemplazando a los medios de comunicación tradicionales como fuentes informativas, surge la interrogante de si la formación de la opinión pública continúa siendo verdaderamente libre. Esta preocupación se intensifica con el avance de la inteligencia artificial, que permite difuminar la verdad y condicionar la libertad de elección mediante *deepfakes*, propaganda computacional y segmentación algorítmica, planteando un nuevo desafío para el cual nuestro sistema electoral podría no estar adecuadamente preparado.

Palabras clave: Libertad de expresión, libertad de información, opinión pública, elecciones, inteligencia artificial, desinformación, *deepfake*, Cambridge Analytica.

Abstract: Democratic electoral processes are grounded in the free formation of public opinion, based on truthful, pluralistic, and verified information, and freely expressed through voting. However, in an environment where it is increasingly difficult to discern between truthful information and disinformation, and where social networks are replacing traditional media as sources of information, the question arises as to whether the formation of public opinion continues to be truly free. This concern is intensified by the advancement of artificial intelligence, which allows the truth to be blurred and the freedom of choice to be conditioned through deepfakes, computational propaganda, and algorithmic segmentation, posing a new challenge for which our electoral system might not be adequately prepared.

Key words: Freedom of expression, freedom of information, public opinion, elections, artificial intelligence, disinformation, deepfake, Cambridge Analytica.

SUMARIO: 1 Introducción. 2 La importancia de la libre formación de la opinión pública. 3 La manipulación del electorado a través de propaganda cognitiva y *microtargeting*: el caso de Cambridge Analytica. 4 El uso de la inteligencia artificial para atacar a los candidatos: el caso de Eslovaquia. 5 Conclusiones. 6 Bibliografía.

1 Introducción

Vivimos en un momento complejo para la información. Hemos pasado de un escenario en el que tan solo unos pocos controlaban la elaboración de noticias y las ponían a disposición del público, primero a través de la prensa impresa y después a través de la radio y la televisión, a otro mucho más dinámico, extenso y con fácil acceso para cualquier persona¹.

Este fácil acceso se manifiesta tanto desde el punto de vista activo como pasivo. Así, desde el punto de vista activo, nos encontramos con que la universalización en el uso de internet y el desarrollo de las redes sociales ha cambiado el propio concepto de informador, al no resultar ya necesario pasar por el filtro de una dirección o una línea editorial para poder publicar, ni siquiera cursar una formación o especialización habilitante, sino que cualquier persona tiene la posibilidad de difundir cualquier

¹ En este sentido, señala el *Informe C: Desinformación en la era digital* de la Oficina de Ciencia y Tecnología del Congreso que, si bien internet y el desarrollo digital propician grandes avances y beneficios económicos y sociales, también han dado pie a “un nuevo contexto social e informativo que ha favorecido una amplificación sin precedentes de la desinformación y sus efectos, convirtiéndola en una destacada amenaza para los sistemas democráticos. Se trata de una cuestión de seguridad nacional que alcanza niveles críticos en situaciones de gran relevancia social, como crisis sanitarias, conflictos bélicos o los procesos electorales”. Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C). Informe C: Desinformación en la era digital. www.doi.org/10.57952/j3p6-9086 (2023) p. 1/26.

contenido y comentar o replicar la información publicada llegando a una audiencia antes impensable fuera de los medios tradicionales².

Esto supone una multiplicación y diversificación del informador debido a la socialización tecnológica de la producción de contenidos y, por tanto, a la democratización de la transmisión de la información³ que hace que cada vez sea más difícil distinguir entre el periodismo profesional y la información partidista, y entre medios de comunicación profesionales y medios alternativos⁴ o *pseudomedios*⁵.

Desde el punto de vista pasivo, nos encontramos con un consumidor de información voraz, que tiene un océano de contenidos al alcance de su mano⁶ al que con gran acierto se refiere Carlos Rodríguez como *catering* de fuentes de información⁷, y que muestra cada vez mayor tendencia al consumo compulsivo de información, otorgando en muchas ocasiones más importancia a la rapidez y a la accesibilidad que a la calidad y a la fiabilidad⁸.

Aquí entra en juego la ley de la oferta y la demanda, provocando que, en un ecosistema en el que el consumidor quiere consultar información sobre un hecho al instante de que este ocurra, algo claramente incompatible con la diligente labor periodística de búsqueda de información y contrastación de la noticia, muchos medios no hayan dudado en plegarse a la voluntad de sus lectores, llegando a arriesgar la veracidad y el rigor de sus publicaciones a cambio de conseguir la primicia, la exclusiva, a cambio de ser los primeros en ofrecer al público una información⁹.

De este modo, el deber de contrastación de la noticia, ese que tradicionalmente había evitado la publicación de informaciones falsas o incluso había salvado la responsabilidad de medios y periodistas por la publicación de información errónea, va

Regap



ESTUDIOS

² ELVIRA PERALES, A., "¿Quién es periodista?", *Teoría y Realidad Constitucional*, 52, 2023, pp. 209-231. <https://doi.org/10.5944/trc.52.2023.39015>, pp. 210-211; BARRIENTOS-BÁEZ, A, CALDEVILLA-DOMÍNGUEZ, D., y YEZERS'KA, L., "Fake news y posverdad: Relación con las redes sociales y fiabilidad de contenidos". *Fonseca, Journal of Communication*, 24, 2022, pp. 149-162. <https://doi.org/10.14201/fjc.28294>. p. 154; MORENO-CABANILLAS, A., CASTILLERO-OSTIO, E., y SERNA-ORTEGA, Á., "El impacto de las redes sociales en la campaña política: Elecciones Generales de 2023 en España", *Redmarka. Revista de Marketing Aplicado*, vol. 28, núm. 1, 2024, pp. 56-76. <https://doi.org/10.17979/redma.2024.28.1.10>, p.58; GALDÁMEZ MORALES, A., "Posverdad y crisis de legitimidad: El creciente impacto de las fake news", *Revista Española de la Transparencia*, vol. 8, 2018, pp. 25-44, <https://doi.org/10.51915/ret.45>, p. 32.

³ GALDÁMEZ MORALES, A., 2018, *op. cit.*, p. 32.

⁴ DEL-FRESNO-GARCÍA, M., "Desórdenes informativos: sobreexpuestos e ininformados en la era de la posverdad", *El profesional de la información*, v. 28, n. 3, e280302. <https://doi.org/10.3145/epi.2019.may.02>, (2019), p. 6.

⁵ Se consideran como *pseudomedios* aquellos sitios web que imitan el estilo y diseño utilizado por los medios y periodistas *mainstream*, pero con fines antagónicos a la ética periodística, mezclando datos y opinión, todo ello con una gran carga ideológica. Se trata de webs que intentan mostrarse como una alternativa a los medios convencionales pero que muchas veces se limitan a absorber el contenido de estos y reformularlo de acuerdo con una ideología concreta: PALAU-SAMPÍO, D., CARRATALÁ, A., "Injecting disinformation into public space: pseudo-media and reality-altering narratives", *Profesional de la información*, v. 31, n. 3, e310312. <https://doi.org/10.3145/epi.2022.may.12> (2022), p.3.

⁶ Según el Digital Report 2024, el número de usuarios de teléfonos móviles ascendía a principios de 2024 a 5,61 billones, un 69,4% de la población mundial, habiendo sufrido un incremento de 138 millones desde comienzos de 2023. Informe disponible en <https://datareportal.com/reports/digital-2024-global-overview-report> [consulta 17 de octubre de 2024].

⁷ RODRÍGUEZ PÉREZ, C., "No diga fake news, di desinformación: una revisión sobre el fenómeno de las noticias falsas y sus implicaciones", *Revista Comunicación*, 40, 2019, pp. 65-74, DOI: <http://dx.doi.org/10.18566/comunica.n40.a05>, p. 66.

⁸ PAUNER CHULVI, C., "Noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red", *Teoría y Realidad Constitucional*, 41, 2018, pp. 297-318. <https://doi.org/10.5944/trc.41.2018.22123>, p 302; MÉNDEZ, L., "El algoritmo sustituye al periodismo", *Cuadernos de Periodistas*, 2021, pp. 65-70, p. 66. Disponible en <https://www.cuadernosdeperiodistas.com/el-algoritmo-sustituye-al-periodismo/>

⁹ PAUNER CHULVI, C., 2018, *op. cit.*, p. 298.

quedando poco a poco relegado a un segundo o tercer plano, eclipsado por la búsqueda de likes, retuits y de páginas vistas¹⁰.

Y es que, si bien es cierto que nuestra Constitución en su artículo 20.1.d) impone el límite de la veracidad a la información al consagrar el derecho a comunicar o recibir libremente información veraz por cualquier medio de difusión, no se trata de una exigencia absoluta, que supondría casi una forma de censura¹¹, sino que lo que se exige es un deber de diligencia en la comprobación de los hechos, evitando que el comunicador actúe con menosprecio a la veracidad o falsedad de lo comunicado, comportándose de manera negligente e irresponsable al transmitir como hechos verdaderos simples rumores carentes de toda constatación o meras invenciones, rumores o insinuaciones¹².

Sin embargo, en el contexto actual, nos encontramos con que no solo no se prima o se valora la veracidad, sino que incluso se está intentando anular el propio concepto de verdad, convirtiéndola en algo subjetivo. Es lo que se ha venido a llamar *posverdad*¹³, definida por la Real Academia Española como la “distorsión deliberada de una realidad, que manipula creencias y emociones con el fin de influir en la opinión pública y en actitudes sociales”.

Todo puede ser verdad o no según el relato que cada uno quiera creer, quedando la verdad relativizada y subordinada a las emociones y a la realidad deseada y totalmente desvinculada de la realidad fáctica¹⁴. Y ello porque, si nada es verdad, difícilmente puede ser algo mentira, quedando el campo abonado para que pueda difundirse la información que más interese en cada momento para condicionar las decisiones de la población en general en la dirección deseada.

Nos adentramos, por tanto, en un contexto de *infocracia* o régimen de la información¹⁵ en el que priman la abundancia de contenidos sobre la veracidad de los mismos,

¹⁰ MENEU-BORJA, M.; SORIA SALVADOR, A.; AGUAR TORRES, J., “La juventud en el ecosistema mediático: una mirada al consumo y la producción informativos de las nuevas generaciones”, *Zer*, 29(56), 2024, pp. 153-172. (<https://doi.org/10.1387/zer.26192>), p. 155.

¹¹ Establece la Sentencia del Tribunal Constitucional 6/1988, de 21 enero, que «las afirmaciones erróneas son inevitables en un debate libre, de tal forma que, de imponerse “la verdad”, como condición para el reconocimiento del derecho, la única garantía de la seguridad jurídica sería el silencio».

¹² Sobre el requisito de la veracidad, sentencias del Tribunal Constitucional 240/1992, de 21 de diciembre, FJ 5.º y 7.º, y 105/1990, de 6 de junio, FJ. 5.º.

¹³ El término *posverdad* fue popularizado por el expresidente estadounidense Donald Trump en 2016, llegando a ser acuñado como palabra de ese año por *Oxford Dictionaries*. No obstante, la primera aparición de la palabra *posverdad* se la debemos al guionista y dramaturgo Steve Tesich, que la utilizó en 1992 en un artículo para la revista *The Nation* sobre el escándalo Watergate y la guerra de Irak, afirmando «Lamento que nosotros, como pueblo libre, hayamos decidido libremente vivir en un mundo en donde reina la *posverdad*».

¹⁴ Afirma Jorge Astudillo que “lo que caracteriza a la *posverdad* es que los hechos objetivos tienen menos credibilidad o menos influencia que los sentimientos, deseos y creencias de las personas al momento de asumir una posición o emitir una opinión sobre un tema de interés público o dar su apoyo o rechazo a una determinada postura social”. ASTUDILLO MUÑOZ, J. L., “Notas sobre la *posverdad*, sus efectos en el sistema democrático y en la protección de los derechos humanos en el marco de la sociedad digital”, *Teoría y Realidad Constitucional*, 52, 2023, pp. 401-427, <https://doi.org/10.5944/trc.52.2023.39023>, p. 404; BARRIENTOS-BÁEZ, A., CALDEVILLA-DOMÍNGUEZ, D. y YEZERS'KA, L., 2022, *op. cit.*, 153.

¹⁵ Sobre el término *infocracia*, señala Byung-Chul Han que la democracia está degenerando en *infocracia* debido a que estamos siendo arrollados por un tsunami de información de la esfera política que provoca distorsiones y trastornos masivos en el proceso democrático debido en gran parte a la digitalización del mundo en que vivimos, que avanza inexorable sometiendo nuestra percepción, nuestra relación con el mundo y nuestra convivencia a un cambio radical. Define este autor el régimen de la información como “la forma de dominio en la que la información y su procesamiento mediante algoritmos e inteligencia

un mal que se hizo especialmente patente durante la pandemia del Covid-19, en la que la propia Organización Mundial de la Salud afirmó que su lucha se libraba tanto contra la pandemia como contra la *infodemia*¹⁶ para referirse a la abundancia tal de información sobre un asunto concreto que resulta difícil discriminar la información verdadera de la mera desinformación¹⁷.

En este escenario, de por sí propicio para que campe la desinformación a sus anchas, irrumpen la inteligencia artificial¹⁸ para revolucionarlo todo, al introducir supuestos que hasta ahora eran impensables y que suponen un reto a la labor de contrastación de los profesionales del periodismo, pero, sobre todo y más importante, supone un gran riesgo para la ciudadanía de a pie, para quien distinguir contenidos veraces de contenidos artificiales se torna más difícil que nunca, por no decir imposible.

Esto cobra especial relevancia en el marco del proceso electoral, un proceso que debe contar con todas las garantías para que la participación de la ciudadanía tenga lugar de manera libre e informada y a través de sufragio libre, igual, directo y secreto y en el que los medios de comunicación jugaban un importante papel, poniendo a disposición del electorado información plural y veraz para una libre formación de la opinión pública.

La importancia de los efectos de la desinformación sobre los procesos electorales se deriva precisamente de la trascendencia que tiene el proceso electoral, mediante el cual la ciudadanía elige a las personas que la van a representar en las distintas instancias durante los próximos cuatro años, sin capacidad de volver atrás esa decisión ni de revocar el poder de representación otorgado, incluso en el caso de que en la formación de su decisión en favor de una u otra opción hubiera intervenido engaño o mentira en forma de desinformación.

Además, en el momento actual, en el que la fragmentación de las opciones electorales y la propia desafección de la ciudadanía hacia la política puede provocar un retraso en la decisión del voto¹⁹, la información o desinformación publicada en los

artificial determinan de modo decisivo los procesos sociales, económicos y políticos". HAN, B., *Infocracia. La digitalización y la crisis de la democracia*, Ed. Taurus, 2022, p. 25. y p. 10.

¹⁶ Así lo señaló el director general de la OMS en la Conferencia de Seguridad de Múnich el 15 de febrero de 2020: "Pero no estamos luchando únicamente contra una epidemia; estamos luchando contra una infodemia. Las noticias falsas se propagan con más rapidez y facilidad que el propio virus, y son igual de peligrosas". Disponible en <https://www.who.int/es/director-general/speeches/detail/munich-security-conference>

¹⁷ Señala David García-Marín que, si bien el término fue popularizado por la OMS, "ya en 2003 este término se utilizó para describir situaciones donde un conjunto de hechos, mezclados con miedo, especulación y rumores, son amplificadas y distribuidos a una audiencia mundial gracias al uso de las tecnologías de la información, con posibilidades de impactar de forma desproporcionada en la seguridad, economía y política de los países", GARCÍA-MARÍN, D., "Infodemia global. Desórdenes informativos, narrativas fake y fact-checking en la crisis de la Covid-19", *Profesional de la información*, v. 29, n. 4, 2020, e290411. <https://doi.org/10.3145/epi.2020.jul.11>.

¹⁸ El Reglamento europeo de inteligencia artificial define un sistema de IA como un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.

¹⁹ WILLOCOQ, S., "Explaining Time of Vote Decision: The Socio-Structural, Attitudinal, and Contextual Determinants of Late Deciding", *Political Studies Review*, 17(1), 2019, pp. 53-64. <https://doi.org/10.1177/1478929917748484>, p. 59.

días previos a la votación resulta más relevante, inclinando definitivamente la balanza del electorado indeciso en favor de una u otra opción.

En un contexto en el que los desórdenes informativos son la norma, ¿realmente somos libres en nuestras decisiones? ¿Podemos seguir hablando de una libre formación de la opinión pública? ¿Podría una potencia extranjera influir en la elección de nuestros máximos representantes? ¿Qué papel puede jugar la inteligencia artificial en todo esto?

2 La importancia de la libre formación de la opinión pública

La libre formación de la opinión pública²⁰ se considera como un elemento intrínseco al sistema democrático, una institución indisolublemente ligada con el pluralismo político, consagrado por el artículo 1 de nuestra Constitución como valor fundamental y requisito del funcionamiento del Estado de derecho, ya que no hay opinión pública sin democracia, ni democracia sin opinión pública²¹.

Considera nuestro Tribunal Constitucional que las libertades del artículo 20 de la Constitución aparecen como una garantía de la opinión pública libre, sin la cual “*quedarían vacíos de contenido real otros derechos que la Constitución consagra, reducidas a formas huecas las instituciones representativas y absolutamente falseado el principio de libertad democrática que enuncia el art. 1 ap. 2º CE y que es la base de nuestra organización jurídico-política*”²².

Como garantía de esta opinión pública libre, el Tribunal señala las libertades del artículo 20 de la Constitución, de modo tal que garantizar las libertades de expresión e información, permitiendo a la población acceder libremente a información plural y veraz, supone también garantizar la libre formación de la opinión pública, pues en base a esta información podrán formarse de manera responsable su propia opinión sobre los distintos asuntos en atención a los hechos²³.

Hoy en día, 40 años después de que el TC escribiera esas palabras, la población tiene más acceso que nunca a la información; de hecho, nunca hemos tenido acceso a tal cantidad de información con tan poco esfuerzo y, sin embargo, no parece que se pueda afirmar que estamos mejor informados, y ello se debe en gran parte a la calidad de la información que tenemos a nuestra disposición, por distintos motivos.

El primero es que, frente al sistema anterior, en el que la información provenía con carácter principal de los medios de comunicación, sujetos a responsabilidad en el caso de que difundieran informaciones falsas sin haber desplegado una diligente

²⁰ Define Antonio Torres del Moral la opinión pública como el precipitado de ideas, creencias y sentimientos que se expresa en la vida pública, ya que, sin expresión externa, la opinión pierde su carácter de pública y deja de cumplir las importantes funciones que la caracterizan. TORRES DEL MORAL, A. (dir.), *Libertades informativas*, Colex, 2009, p. 140.

²¹ TORRES DEL MORAL, A. (dir.), 2009, *op. cit.*, p. 117.

²² Sentencia del Tribunal Constitucional 12/1982, de 31 de marzo, FJ 3.º.

²³ TORRES DEL MORAL, A. (dir.), 2009, *op. cit.*, p. 142.

labor de contrastación, a día de hoy cualquiera puede ser comunicador²⁴, cualquiera puede subir informaciones a la red en una suerte de “periodismo ciudadano”²⁵, sin más límites que los establecidos para la libertad de expresión, que necesariamente deben ser mínimos en un sistema democrático²⁶.

Esta proliferación de comunicadores no habría sido posible sin el desarrollo de las nuevas tecnologías y el ascenso de las redes sociales y de mensajería como principales formas de acceso a la información y a las noticias²⁷. Hoy en día, gran parte de la población no se informa a través de los medios tradicionales, ni siquiera de la prensa digital, sino que lo hacen a través de su *feed* de Facebook, X, o la red social que cada uno siga²⁸, y que los medios de comunicación utilizan como altavoz para sus publicaciones²⁹, a menudo haciendo uso de titulares sensacionalistas o *clickbait*³⁰, que podemos considerar desinformación en sí mismos ya que no aportan nada al periodismo, solo a la cuenta de resultados.

Es la red y su algoritmo la que decide qué contenido muestra a cada persona usuaria, controlando por tanto el flujo de la información, sin que la ciudadanía pueda conocer los sesgos ideológicos o intereses detrás de la red debido a la falta de transparencia de sus algoritmos³¹.

El ascenso de las redes sociales como propagadores de noticias se ha visto impulsado por distintos motivos, entre los que destacan la inmediatez en la difusión, el acceso simultáneo a una gran audiencia, la accesibilidad de las redes a cualquier persona (tanto para publicar como para recibir contenidos) y, aunque muchas veces las propias personas usuarias no sean conscientes, la facilidad para establecer filtros y segmentar con gran detalle usuarios para fines concretos³².

Regap



ESTUDIOS

²⁴ Así lo señala el Tribunal Constitucional en la Sentencia 8/2022, de 27 de enero, FJ 2.º, cuando dice que “*el uso de las herramientas digitales convierte a sus usuarios en creadores de contenidos, emisores, difusores y reproductores de esos contenidos (...). Por tanto, los usuarios pueden llegar a desempeñar un papel muy cercano al que venían desarrollando hasta ahora los periodistas en los medios de comunicación tradicionales*”.

²⁵ ASTUDILLO MUÑOZ, J. L., 2023, *op. cit.*, p. 409.

²⁶ ABA CATOIRA, A., “Libertades de expresión e información en la sociedad digital como garantías de la democracia”, *Revista Doctrina Distrital*, vol. 1, n.º extra 2, 2021, pp. 55-70, p. 56; ABA CATOIRA, A., “Los desórdenes informativos en un sistema de comunicación democrático”, *Revista de Derecho Política*, 1(109), 2020, pp. 119-151. <https://doi.org/10.5944/rdp.109.2020.29056>; CARMONA SALGADO, C., *Libertad de expresión e información y sus límites*. EDESA Editoriales de Derecho Reunidas, 1991, p. 69.

²⁷ Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 7/26.

²⁸ Sobre esto profundiza Emily Bell en su artículo “*Facebook se está comiendo el mundo*”. *Cuadernos de Periodistas*, 32, 2016, pp. 67-76. Disponible en <https://www.cuadernosdeperiodistas.com/facebook-se-esta-comiendo-mundo/>

²⁹ JIMÉNEZ PALOMARES, P. y GOMES-FRANCO E SILVA, F., “Visibilidad de la información en redes sociales: los algoritmos de Facebook y su influencia en el *clickbait*”, *Caleidoscopio - Revista Semestral de Ciencias Sociales y Humanidades*, 23(41), 2019, pp. 173-211. doi:10.33064/41crscsh1772C, p. 175.

³⁰ Siguiendo a Berta García y otros, podemos definir el *clickbait* como “el titular de una noticia que no responde a los criterios periodísticos tradicionales y cuyo objetivo final es mantener al receptor en la página el mayor tiempo posible, no informar. En concreto, el titular cebo tendría como objetivo principal la comercialización o difusión de la información, mientras que el titular periodístico mantendría el rol principal de informar a los usuarios”. GARCÍA OROSA, B., GALLUR SANTORUN, S. y LÓPEZ GARCÍA, X., “El uso del *clickbait* en cibermedios de los 28 países de la Unión Europea”, *Revista Latina de Comunicación Social*, 72, 2017, pp. 1.261-1.277. <http://www.revistalatinacs.org/072paper/1218/68es.html>, p. 1270.

³¹ PAUNER CHULVI, C., 2018, *op. cit.*, p. 303.

³² Resulta especialmente relevante la microsegmentación, una práctica publicitaria que permite dirigir contenidos a grupos específicos en base a sus intereses, características, etc., personalizando los mensajes que reciben. Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 14/26.

Esta segmentación hace que recibamos información “cocinada” algorítmicamente en vez de en atención a la relevancia de la noticia o incluso a una determinada línea editorial, lo que puede afectar a la calidad de la propia información³³ y, lo que es aún más importante, al proceso de formación de la opinión pública, ya que esta segmentación o microsegmentación puede ser maliciosamente utilizada con fines desinformativos.

En segundo lugar, porque en este entorno en el que cualquiera puede publicar contenidos y replicarlos sin contrastar lo más mínimo su veracidad, y en el que podemos estar conectados a cientos de personas a través de las redes sociales, se intensifican los efectos de los sesgos cognitivos³⁴, especialmente del sesgo de confirmación, definido como la tendencia de una persona a favorecer y priorizar la información que confirma sus creencias o hipótesis.

Es decir, de entre la cantidad ingente de información e informadores a los que podemos acceder, tenderemos a seleccionar o priorizar las opiniones de aquellos que confirmen las nuestras, creando una suerte de “burbujas de filtro”³⁵ o “cámaras de eco o resonancia”³⁶ en las que el propio grupo cada vez está más convencido de estar en lo cierto, pues prácticamente toda la información que recibe refuerza sus creencias y rechaza las contrarias³⁷.

Ocurre además que para nuestro cerebro resulta muy difícil convivir con incongruencias, por lo que tratamos de rellenar cuanto antes los huecos que quedan en nuestros razonamientos, incluso aunque sea con información falsa³⁸. Preferimos autoengañarnos y disfrutar de la armonía de la coherencia entre nuestros actos y creencias (o entre nuestras creencias entre sí) a tener que enfrentarnos a la ansiedad que supone convivir con dos creencias en conflicto o con el conflicto entre nuestros pensamientos y nuestro comportamiento, lo que Leon Festinger bautizó como *disonancia cognitiva*³⁹.

³³ Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 8/26.

³⁴ Sobre los sesgos cognitivos, GARRIGUES WALKER, A. y GONZÁLEZ DE LA GARZA, L. M., *El Derecho a no ser engañado. Y cómo nos engañan y nos autoengañamos*, Aranzadi, 2020, pp. 85 y ss., siguiendo a Stuart Sutherland, definen el sesgo cognitivo como “un fenómeno psicológico de naturaleza involuntaria, ya que se produce de forma automática, y que describe una alteración en la mente humana que la conduce a una distorsión de la percepción, a una distorsión cognitiva y o a juicios imprecisos o a interpretaciones ilógicas o irracionales de los hechos”.

³⁵ Las burbujas de filtro son un fenómeno por el que una persona se encuentra expuesta principalmente a información y perspectivas que refuerzan sus creencias y puntos de vista preexistentes, debido a algoritmos de recomendación y personalización en plataformas digitales y redes sociales, todo ello sin que se requiera un papel activo por parte del usuario. Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 13/26.

³⁶ Las cámaras de resonancia son un entorno o plataforma donde las ideas, mensajes o conceptos encuentran una audiencia receptiva y amplifican su impacto, requiriendo de un papel activo del usuario, que demanda activamente determinados contenidos por diferentes razones, incluidas las ideológicas, y donde puede circular más libremente la desinformación. Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 13/26.

³⁷ MARTÍN, M., FIOR, A., COSTA ESCUREDO, Á., RUIZ SAN ROMÁN, J. A., “La figura del periodista frente al fenómeno de la desinformación en la era digital”, *Zer*, 29(56), 2024, pp. 65-85. (<https://doi.org/10.1387/zer.25993>), p. 68; BARRIENTOS-BÁEZ, A., CALDEVILLA-DOMÍNGUEZ, D. y YEZERS'KA, L., 2022, *op. cit.*, p.155; Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C) (2023). *Op. cit.*, p. 13/26; BERROCAL-GONZALO, S., WAISBORD, S. y GÓMEZ-GARCÍA, S., “Polarización política y medios de comunicación, su impacto en la democracia y la sociedad”, *Profesional de la información*, v. 32, n. 6, e320622. <https://doi.org/10.3145/epi.2023.nov.22>, 2023, p. 3.

³⁸ GARRIGUES WALKER, A. y GONZÁLEZ DE LA GARZA, L. M., 2020, *op. cit.*, p. 64.

³⁹ Leon Festinger considera que la disonancia, la existencia de relaciones entre conocimientos, opiniones, creencias o conductas que no concuerdan, es un factor de motivación, ya que condiciona nuestra conducta en el sentido de dirigirla a reducir

En tercer lugar, y aquí reside probablemente la mayor amenaza, por la irrupción de las herramientas de inteligencia artificial que permiten, por un lado, un uso mucho más preciso, eficiente y masivo de los datos de los electores, como ocurrió en las elecciones presidenciales estadounidenses de 2016 con Cambridge Analytica. Por otro lado, nos permiten construir una realidad a medida, fabricando los datos, imágenes, vídeos o audios necesarios para sustentar esa creación y haciéndolo de tal forma que resulten indistinguibles de la propia realidad, lo que se conoce como *deepfakes*⁴⁰.

Uno de los problemas clave de las *deepfakes* es que atacan a los propios principios de la estructura social en la que la mayoría de la sociedad asume que hay alguna forma de verdad mutuamente aceptada, como eran hasta ahora los contenidos audiovisuales, por lo que la irrupción de una tecnología que rompe con ese acuerdo social o, al menos, lo pone en entredicho, supone una pérdida de confianza epistémica⁴¹ que deriva en el fin de la creencia en la verdad⁴². Y esto supone no solo que podamos tomar por cierto un contenido fabricado, sino también que podamos tener por falso un audio, vídeo o imagen real⁴³.

Por tanto, nos encontramos intentando sobrevivir en un océano de información, en su mayoría no contrastada, en el que convive información veraz con información tergiversada o sacada de contexto, parodias, información falsa, información atribuida a fuentes incorrectas, información fabricada... y es basándonos en aquello que seleccionemos de este inmenso océano cómo formaremos nuestra opinión sobre los distintos temas y, por tanto, cómo se configurará la opinión pública.

En este escenario, para cualquiera con los medios adecuados resultaría relativamente sencillo influir en la formación de la opinión pública, orientando a los electores en una dirección u otra alterando el sentido de su voto e influyendo, por tanto, de manera clara en el resultado de los procesos electorales, con las implicaciones que ello puede tener incluso a nivel de seguridad nacional, especialmente cuando quien maneja los hilos es una potencia extranjera⁴⁴.

Para intentar controlar este riesgo, la Unión Europea creó dos comisiones sucesivas de Injerencias Extranjeras, la primera en 2020 y la segunda en 2022. Fruto del

esa disonancia. Extractado por A OVEJERO a partir de la obra de L. Festinger: Teoría de la disonancia cognoscitiva. *La teoría de la disonancia cognoscitiva*. Psicothema, 5(1), 1993, pp. 201-206, Recuperado a partir de <https://reunido.uniovi.es/index.php/PST/article/view/7156> p. 202.

⁴⁰ Si bien los primeros usos de la tecnología *deepfake* se encuentran en el ámbito de la pornografía, la primera documentación de esta tecnología la encontramos en el año 2017, cuando tres investigadores de la Universidad de Washington demostraron cómo se podía manipular la tecnología de edición de imágenes dinámicas para crear nuevos contenidos, concretamente imitando la voz y expresiones faciales del presidente Obama. SUPASORN SUWAJANAKORN, STEVEN M. SEITZ, AND IRA KEMELMACHER-SHLIZERMAN. *Synthesizing Obama: Learning Lip Sync from Audio*. ACM Trans. Graph. 36, 4, Article 95 (July 2017), 13 pages. (2017) DOI: <http://dx.doi.org/10.1145/3072959.3073640>.

⁴¹ LEÓN MENDOZA, R., "La imagen como forma de (des)conocimiento en la era del deepfake", *ANIAV Revista de Investigación en Artes Visuales*, n. 11, 2022, pp. 53-70, <https://doi.org/10.4995/aniav.2022.17309>. p. 3.

⁴² GONZÁLEZ DE LA GARZA, L. M., "Por qué las campañas electorales cognitivas basadas en la posverdad pueden erosionar las democracias de opinión", *Revista General de Derecho Constitucional*, 32 2020, p. 9.

⁴³ TERNOVSKI, J., KALLA, J., y ARONOW, P. M., *Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discrepancy: Evidence from Two Experiments*. <https://doi.org/10.31219/osf.io/dta97>, 2021, p. 6.

⁴⁴ Según los datos del Eurobarómetro Flash n.º 522 (2023) sobre "Democracia", el 43 % de los europeos mencionó como una de las amenazas más graves para la democracia la propaganda o la información falsa o engañosa procedente de fuentes extranjeras no democráticas y la injerencia extranjera encubierta en la política y la economía de su país.

trabajo de estas comisiones, el 1 de junio de 2023 el Parlamento Europeo aprobó la Resolución P9_TA(2023)0219 sobre las injerencias extranjeras en todos los procesos democráticos de la Unión Europea, en particular la desinformación.

En esta resolución, que busca acordar una estrategia común para la Unión en la lucha contra las injerencias extranjeras, se señala que estas pueden adoptar diversas formas: desde fraude en las papeletas o bloqueo de los colegios electorales hasta la distribución de información distorsionada sobre los candidatos o el propio proceso electoral y campañas de desinformación⁴⁵, pasando por financiación a determinadas opciones políticas o apoyo a corrientes ideológicas extremas.

En opinión del Parlamento, el objetivo de estas injerencias es distorsionar la política nacional para socavar o destruir la confianza a largo plazo de la ciudadanía en la legitimidad de sus instituciones y procesos democráticos⁴⁶, constituyendo no solo una amenaza a la libre formación de la opinión pública, sino también a la seguridad nacional y transfronteriza⁴⁷.

3 La manipulación del electorado a través de propaganda cognitiva y *microtargeting*: el caso de Cambridge Analytica

La propaganda cognitiva o marketing cognitivo es un tipo de publicidad que se centra en aprovechar los sentimientos o la emoción para conseguir una reacción en la persona destinataria. Se estudia al usuario a fondo, sus gustos, los contenidos que publica, lo que comparte, buscando alinear la marca en cuestión con todo aquello que le gusta al consumidor o que este considera positivo o deseable, de tal forma que se sienta “naturalmente” atraído hacia ella. Lo novedoso es que no es una propaganda pasiva, sino que se aprovecha de los sesgos caracterológicos de las personas destinatarias para diseñar una campaña de gran precisión a la medida de cada persona y de sus preferencias emocionales y políticas⁴⁸.

Antes de la expansión de las redes sociales este tipo de publicidad resultaba casi impensable, ya que por grande que fuera el esfuerzo de conocer al usuario a través de encuestas o del estudio de sus compras, nunca ha habido tanta facilidad como ahora para acceder a toda esa información y procesarla.

⁴⁵ Resolución P9_TA(2023)0219 del Parlamento Europeo, de 1 de junio de 2023, sobre las injerencias extranjeras en todos los procesos democráticos de la Unión Europea, en particular la desinformación (2022/2075(INI)), Considerando AZ.

⁴⁶ Resolución P9_TA(2023)0219 del Parlamento Europeo, de 1 de junio de 2023, sobre las injerencias extranjeras en todos los procesos democráticos de la Unión Europea, en particular la desinformación (2022/2075(INI)), Considerando BA.

⁴⁷ La Agencia de la Unión Europea para la ciberseguridad (ENISA) y el Servicio Europeo de Acción Exterior (EEAS) hablan de injerencia y manipulación extranjera de la información (FIMI, por sus siglas en inglés) para referirse a un patrón de comportamientos, principalmente no ilegales que amenaza o tiene el potencial de impactar negativamente en los valores, procedimientos y procesos políticos. Se trata de una actividad manipulativa por naturaleza, llevada a cabo de manera intencionada y coordinada, cuyos actores pueden ser tanto actores estatales como no estatales, incluidos sus representantes dentro y fuera de su propio territorio. ENISA, EEAS. *Foreign Information Manipulation Interference (FIMI) and Cybersecurity - Threat Landscape*. DOI: 10.2824/750. (2022) Disponible en https://www.cde.ual.es/wp-content/uploads/2022/12/EEAS-ENISA-Disinformation_Misinformation.pdf.

⁴⁸ GONZÁLEZ DE LA GARZA, L. M., 2020, *op. cit.*, p. 20.

El mercado tecnológico nos ofrece multitud de aplicaciones y servicios supuestamente gratuitos que instalamos en nuestros dispositivos y comenzamos a usar facilitando todos los datos que el servicio en cuestión solicita para configurar nuestro perfil y así mejorar la experiencia de usuario mediante la personalización de los contenidos. Todo ello sin ser conscientes o menospreciando el hecho de que, como ocurre en otros ámbitos, si no nos cobran por el producto, es que el producto somos nosotros⁴⁹.

En la red queda registro de nuestro perfil, edad, sexo, ubicación, intereses declarados... pero también de cada una de nuestras acciones e interacciones dentro de la propia red: qué es lo que publicamos, qué cuentas o páginas seguimos, qué contenidos nos gustan o replicamos, incluso cual es nuestra reacción ante determinadas publicaciones, anuncios, noticias... Queda constancia de cuales son nuestros medios de comunicación o tendencias políticas afines a través de las noticias que publicamos, comentamos o replicamos, pero también a través de los políticos que seguimos o criticamos... en resumen, toda nuestra actividad a través de las redes sociales queda registrada y perfectamente segmentada por sexo, ideología, ubicación, franja de edad, etc., para ser usada de la forma que más convenga al mejor postor.

Precisamente esto es lo que ocurrió en 2016 durante las elecciones presidenciales estadounidenses, en las que la empresa Cambridge Analytica, una rama del Grupo SCL⁵⁰, contratada por el candidato Donald Trump, utilizó indebidamente datos personales de millones de usuarios de Facebook para crear perfiles psicológicos concretos de los usuarios y así diseñar anuncios políticos con el tono y contenido adecuados para cada perfil con el objetivo de dirigir su sentido del voto, a la vez que se creaban y difundían noticias falsas que aumentaban el efecto buscado. Donald Trump ganó las elecciones y no podemos saber cuál habría sido el resultado de no haber recurrido a este tipo de prácticas.

No estamos hablando de propaganda electoral al uso, sino de publicidad hecha a la medida de cada usuario utilizando perfiles psicométricos, que busca remover sus emociones para conseguir una actuación por su parte⁵¹. El objetivo ya no es convencer al usuario para que vote a un candidato determinado, sino generar en él la emoción que provoca que se convenza solo. Por tanto, la propaganda cognitiva no apela a la razón, sino a las emociones, a la alegría, a la rabia, a la ilusión y la venganza... de tal modo que el propio destinatario no llega a ser consciente de que está siendo influenciado, de que verdaderamente no está decidiendo libremente.

⁴⁹ Sobre este tema: MARTÍN-HERRERA, I. y GUERRERO-CABALLERO, M., "Millennials y Generación X frente a la realidad del Big data y la protección de datos personales en Internet", *Vivat Academia*, 157, 2024, pp. 1-20. <http://doi.org/10.15178/va.2024.157.e1489>.

⁵⁰ SCL Group (Strategic Communication Laboratories) es una empresa británica fundada en 1990, dedicada a la investigación del comportamiento y comunicación estratégica.

⁵¹ Resulta muy didáctica la explicación de BYUNG-CHUL HAN sobre los efectos del *microtargeting* en su libro *Infocracia*: "En el *microtargeting*, los votantes no están informados del programa político de un partido, sino que se les manipula con publicidad electoral adaptada a su psicograma, y no pocas veces con *fake news*. (...) Los ciudadanos dejan de estar sensibilizados para las cuestiones importantes, de relevancia social. (...) De este modo, socavan un principio fundamental de la democracia: *la autoobservación de la sociedad*". HAN, B., 2022, *op. cit.*, p. 37.

Para conseguir los datos, recurrieron a un académico de la Universidad de Cambridge, Aleksandr Kogan, que ideó una aplicación *online* llamada *This is your digital life* que se presentaba como un test online y afirmaba ser una herramienta de investigación, que incluso remuneraba simbólicamente a las personas participantes. Para finalizar el test, los usuarios debían iniciar sesión en Facebook y otorgarle ciertos permisos a la aplicación, como ocurre en tantos otros casos⁵².

Aproximadamente 270 mil personas rellenaron la encuesta y concedieron los permisos que la aplicación solicitaba, creyendo que sus datos solo serían utilizados en el seno de la investigación y con fines académicos. Sin embargo, entre los permisos que la aplicación solicitaba estaba el acceso a los datos de los amigos en la red social, lo que permitió a la empresa acceder a los datos de unos 87 millones de personas⁵³.

Conseguidos los datos, Cambridge Analytica se sirvió de herramientas de *big data*⁵⁴ y *machine learning*⁵⁵ para tratarlos en masa, cruzando las respuestas de las encuestas con los *likes* y publicaciones compartidas en Facebook por los distintos usuarios. Esto les permitió segmentar a los 87 millones de usuarios en pequeños grupos de personas o perfiles psicográficos⁵⁶ preocupadas por asuntos concretos y que, por tanto, serían receptivas a publicaciones sobre esas materias concretas si se utilizaba el tono adecuado.

Esta minería de datos, que con medios tradicionales habría sido imposible, se convierte en algo sencillo y rápido gracias a la inteligencia artificial y, concretamente, al *machine learning*, que permite obtener en tiempo récord una segmentación extremadamente escrupulosa y detallada.

Una vez segmentados los votantes en las distintas categorías, se utilizó un sistema de inteligencia artificial capaz de analizar cientos o incluso miles de variaciones de un anuncio para comprobar cuál era la que mejor podría funcionar con cada uno de los perfiles psicográficos previamente identificados. Se trataba de personalizar hasta el más mínimo detalle de cada anuncio, desde el tipo y color de letra hasta el tono del mensaje, pasando por la iconografía o la música, en el caso del contenido audiovisual.

Por tanto, a cada pequeño grupo se mostraba solo la variación del anuncio que el sistema había detectado que mejor funcionaría, llegando a diseñarse entre 40 y 50

⁵² ISAAK, J. y HANNA, M. J., "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection", in *Computer*, vol. 51, n.º 8, 2018, pp. 56-59, doi: 10.1109/MC.2018.3191268, p. 57.

⁵³ En un primer momento se pensó que Cambridge Analytica había accedido a los datos de aproximadamente 50 millones de usuarios, pero los datos a fecha 1 de mayo de 2018 proporcionados por Meta, propietaria de Facebook, apuntan a 87 millones de personas, la mayoría de ellas en Estados Unidos (81,6%). Datos disponibles en la web: <https://about.fb.com/news/2018/04/restricting-data-access/>

⁵⁴ El analista Dan Kusnetzky define *big data* como las herramientas, procesos y procedimientos que permiten a una organización crear, manipular y administrar grandes conjuntos de datos e instalaciones de almacenamiento. CAMARGO ORTEGA, J. F., CAMARGO VEGA, J. J., y JOYANES AGUILAR, L., "Conociendo Big Data", *Revista FI-UPTC*, 24(38), 2015, pp. 63-77, p. 63.

⁵⁵ El *machine learning* o aprendizaje automático es una rama de la inteligencia artificial que permite a las computadoras aprender de los datos sin ser programadas explícitamente para cada tarea. En lugar de seguir instrucciones fijas, los sistemas de *machine learning* analizan patrones en los datos, lo que les permite hacer predicciones o tomar decisiones basadas en la información que han procesado.

⁵⁶ Se define la psicografía como el estudio y clasificación de las personas según sus actitudes, aspiraciones y otros criterios psicológicos.

mil variantes de diferentes argumentos electorales⁵⁷, que se adaptaban a la perfección al perfil del usuario al que se mostraban, sin que además pudieran ser conocidos por otros usuarios, ya que se basan en publicaciones invisibles o *dark post*⁵⁸, lo que hace que también sean difícilmente rastreables por los organismos de control.

Es decir, un usuario no puede ver este tipo de publicaciones en el muro de otro usuario, sino que solo se muestran a su destinatario, aumentando así la personalización y evitando que el usuario se percate de que está siendo manipulado, al tiempo que se evita o retrasa un debate público y abierto sobre el contenido y la veracidad de esas publicaciones que permitiría contrastarlas⁵⁹.

Además de todo lo expuesto, debemos tener en cuenta que la propaganda cognitiva no opera de manera aislada. No se trata solo de que se muestren al usuario anuncios especialmente diseñados para él en función de sus intereses, previamente recabados a través de las cesiones de datos efectuadas, consciente de ello o no, por el propio usuario, sino que a menudo se combina esta herramienta con otras como el uso de cuentas falsas automatizadas⁶⁰ en redes sociales, también conocidas como *social bots*⁶¹.

En el caso de las elecciones presidenciales estadounidenses de 2016, el uso de cuentas automatizadas fue estratégico y deliberado, especialmente en el lado republicano, ajustando a la perfección el momento de producción de contenido durante los debates y llegando en el día de las elecciones a superar los *bots* pro-Trump a los de Clinton a razón de 5 a 1⁶².

Estas cuentas automatizadas se utilizan sobre todo para alterar el debate en la propia red, dando audiencia a publicaciones que podrían parecer inicialmente irrelevantes, llegando a viralizar contenidos que de otro modo habrían pasado desapercibidos, apropiándose de un debate creado inicialmente por humanos o incluso creando *ex novo* un debate que de otra forma no existiría⁶³. Todo ello con un volumen de publicaciones e interacciones muy superior al humanamente posible, lo que consigue con poco esfuerzo manipular el peso del apoyo a las distintas opciones,

Regap



ESTUDIOS

⁵⁷ GARRIGUES WALKER, A. y GONZÁLEZ DE LA GARZA, L. M., *op. cit.*, p. 57.

⁵⁸ Son publicaciones cuya visibilidad se condiciona a la pertenencia a determinados grupos seleccionados por el administrador de la página. Se trata de una funcionalidad especialmente pensada para el mundo publicitario, que permite no saturar al mercado mostrando información que pueda no resultar relevante para grupos determinados, por ejemplo por edad o ubicación, y dirigirse solamente a la audiencia a la que la publicación puede interesar.

⁵⁹ ANDERSON, B. y HORVARTH, B., PÉREZ LAGÜELA, Elena (trad.). "El ascenso de la máquina de armas propagandísticas de inteligencia artificial", *Papeles de relaciones ecosociales y cambio global*, 38, 2017, pp. 37-47, p. 40.

⁶⁰ Una cuenta automatizada o *bot* es un software o programa informático elaborado para llevar a cabo determinados tipos de actividades repetitivas como si de un ser humano se tratase; puede generar contenido, replicarlo, seguir o dejar de seguir cuentas, enviar mensajes directos... todo ello imitando el comportamiento humano. SANTOS SANZ, L., "Identificación automática de bots en Twitter basada en contexto", *Reports de inteligencia económica y relaciones internacionales*, n.º 13 (Vol.), 2023, Disponible en <http://hdl.handle.net/10486/700265>, p. 5.

⁶¹ Un *social bot* es una cuenta controlada por software que participa activamente en las redes sociales para influir en la opinión pública en la dirección deseada: PASTOR-GALINDO, J. *et al.*, "Twitter social bots: The 2019 Spanish general election data", *Data in Brief*, Volume 32, 2020, <https://doi.org/10.1016/j.dib.2020.106047>.

⁶² ANDERSON, B. y HORVARTH, B., PÉREZ LAGÜELA, E. (trad.), 2017, *op. cit.*, p. 43.

⁶³ LEVI, S., *#FakeYou: Fake news y desinformación*, Rayo Verde Editorial, 2020, p. 82.

inclinando de manera artificial la balanza hacia una de ellas que aparece claramente como ganadora⁶⁴.

Todo esto consigue reforzar la burbuja del usuario y con ella la polarización existente, una polarización en la que los extremos ya no son tan mal vistos⁶⁵, pues hay gran cantidad de cuentas que defienden sus posiciones al calor del anonimato de las redes y que, además, encuentran explicación a lo inexplicable, normalmente en base a alguna retorcida conspiración que casualmente lo explica todo⁶⁶.

Todo ello amplificado por el efecto de la disonancia cognitiva: necesitamos buscar explicaciones que refuercen nuestras creencias y esto hace casi imposible, a medida que se avanza en esta madriguera, dar marcha atrás, rectificar nuestros pensamientos, pues preferiremos creer cualquier cosa antes que rechazar todas las teorías que ya hemos defendido y que, al menos en el contexto de nuestra burbuja, encajan a la perfección. La credibilidad de la información es directamente proporcional a cuanto coincida con nuestros valores, creencias y prejuicios⁶⁷.

Resulta importante, no obstante, aclarar que el problema no es el uso de las redes sociales para llegar al electorado –de hecho, la campaña de otro candidato a la presidencia de Estados Unidos, Barack Obama, fue de las primeras en explotar las redes sociales, en concreto Facebook y MySpace⁶⁸ para hacer llegar su mensaje–, la clave es que lo que hacía era explotar las herramientas que la red social ponía a su servicio para llegar al electorado a través de esa vía⁶⁹.

Por el contrario, en el caso de la campaña de Donald Trump, lo que realmente hizo Cambridge Analytica fue engañar a los usuarios para que les confiaran sus datos, su información, e utilizarla con fines totalmente distintos a los declarados. Esto supone una clara violación de la debida protección de datos de las personas usuarias, que resulta aún de mayor gravedad porque se produce con fines electorales.

El objetivo es modificar el sentido del voto de los usuarios de la red escudriñando en su privacidad para utilizar los datos robados con el objetivo de hacer creer a ese usuario que es él quien está eligiendo, cuando realmente lo que ocurre es que se ha condicionado su voluntad a un nivel profundo, mostrando solo los contenidos que reafirmaban la postura que interesaba y creando propaganda a la medida del usuario aprovechando la información robada.

⁶⁴ Se calcula que en la campaña española de las elecciones generales de noviembre de 2019 había más de 40.000 *bots* tuiteando sobre eventos destacados de la campaña. PASTOR-GALINDO, J. *et al.*, "Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election", *IEEE Transactions on Network and Service Management*, vol. 17, n.º 4, 2020, pp. 2156-2170, doi: 10.1109/TNSM.2020.3031573, p. 2158.

⁶⁵ Sobre la polarización: BERROCAL-GONZALO, S., WAISBORD, S. y GÓMEZ-GARCÍA, S., 2023, *op. cit.*

⁶⁶ SÁDABA RODRÍGUEZ, I. y GARCÍA ARNAU, A., "Teorías de la conspiración y crisis globales: retos para la teoría sociológica", *RES. Revista Española de Sociología*, 33(2), 2024, <https://doi.org/10.22325/fes/res.2024.222>, p. 4.

⁶⁷ BARRIENTOS-BÁEZ, A., CALDEVILLA-DOMÍNGUEZ, D. y YEZERS'KA, L., 2022, *op. cit.*, p. 158.

⁶⁸ Al celebrarse las elecciones, Obama reunió 3 millones de seguidores en Facebook, 700.000 en MySpace, 100.000 en Twitter y 144.615 suscriptores en YouTube. RUBIO NÚÑEZ, R., "Quiero ser como Obama (me pido una red social)", *Cuadernos de pensamiento político (Madrid, Spain)*, 21, 2009, pp. 123-154; p. 131.

⁶⁹ LINARES RODRÍGUEZ, V. y ABEJÓN MENDOZA, P., "De la marca Obama 2.0 al antagonismo digital de Donald Trump", *Investigación y desarrollo*, 29(1), 2021, pp. 99-132. <https://doi.org/10.14482/indes.29.1.658.8>, p. 109.

La propaganda computacional se sitúa, por tanto, como una poderosa herramienta contra la democracia⁷⁰, al afectar al sentido de voto a través de la manipulación de la formación de la opinión pública⁷¹.

Evidentemente, puede que el resultado de las elecciones estadounidenses de 2016 hubiese sido el mismo sin la actuación de Cambridge Analytica, pero es imposible saberlo. Lo que sí sabemos es que uno de los candidatos presidenciales llevó a cabo una campaña poco transparente, robando y tratando los datos de millones de personas y utilizando propaganda cognitiva para asegurarse su voto y que esta situación acabó suponiendo la quiebra de Cambridge Analytica y grandes pérdidas para la red social Facebook, mientras que el candidato que contrató los servicios de la consultora y ganó las elecciones siguió en su puesto hasta terminar su mandato.

4 El uso de la inteligencia artificial para atacar a los candidatos: el caso de Eslovaquia

Las herramientas que ofrece la inteligencia artificial, concretamente en el campo de la desinformación, suponen un reto nunca visto. En el apartado anterior se ha profundizado en alguno de los usos de las herramientas de *machine learning* en este campo, pero no podemos terminar sin tratar el papel de las herramientas *deep learning*⁷², que permiten crear contenidos audiovisuales indistinguibles de la realidad, hasta el punto de poder utilizar la imagen o la voz de una persona real para que haga o diga cosas que jamás ha hecho o dicho.

Podemos imaginarnos infinidad de formas en las que esta tecnología puede afectar al proceso de toma de decisiones en materia electoral, pero lo cierto es que ya tenemos ejemplos reales.

El 30 de septiembre de 2023 se celebraron elecciones parlamentarias anticipadas en Eslovaquia en las que el ex primer ministro prorruso Robert Fico, líder de Dirección-Socialdemocracia Eslovaca (SMER-SSD)⁷³, y Michal Šimečka, líder del partido liberal Eslovaquia Progresista (EP), se disputaban el puesto de primer ministro.

Apenas dos días antes de la celebración de los comicios, se publicó en Facebook un audio de unos dos minutos en el que supuestamente se escuchaba al candidato de EP discutiendo con la periodista Monika Tódová la forma de amañar el proceso electoral

⁷⁰ WOOLLEY, S. C. y HOWARD, P. N., "Computational Propaganda Worldwide: Executive Summary", *Working Paper 2017.11*. Oxford, UK: Project on Computational Propaganda. demtech.oii.ox.ac.uk. 14 pp., p. 7.

⁷¹ GONZÁLEZ DE LA GARZA, L. M., 2020, *op. cit.*, p. 22.

⁷² Define Raúl López Briega el *deep learning* o aprendizaje profundo como un subcampo dentro del *machine learning* que utiliza distintas estructuras de redes neuronales para lograr el aprendizaje de sucesivas capas de representaciones cada vez más significativas de los datos. El aprendizaje profundo o *deep learning* hace referencia a la cantidad de capas de representaciones que se utilizan en el modelo; en general se suelen utilizar decenas o incluso cientos de capas de representación que aprenden automáticamente a medida que el modelo es entrenado con los datos. Citado en GUAÑA-MOYA, J., SALGADO-REYES, N., y ESCOBAR-TERÁN, C., "La informática y la ciencia de datos utilizando Deep learning. [Computing and data science using deep learning]", *Revista Ibérica De Sistemas e Tecnologías De Informação*, 2022, pp. 116-127. Disponible en <https://www.proquest.com/scholarly-journals/la-informática-y-ciencia-de-datos-utilizando-deep/docview/2714752347/se-2>.

⁷³ TEJERO MATOS, P., "¿Populismo híbrido? Análisis del SMER-SD en Eslovaquia", *Revista Jurídica de La Universidad Autónoma de Madrid*, 44, 2021, pp. 175-198, p. 186. <https://doi.org/10.15366/rjuam2021.44.006>.

comprando votos de la minoría romaní. Para no dejar lugar a dudas, el audio estaba subtulado e iba acompañado de la imagen estática del candidato liberal y de la periodista. Ambos salieron rápidamente a desmentir el audio, si bien con cierta tibieza, dado que el país se encontraba en el periodo de reflexión, que en el caso eslovaco se extiende durante las 48 horas previas al comienzo de la votación⁷⁴.

También el departamento de comprobación de información de la agencia de noticias AFP⁷⁵ salió rápidamente a poner de manifiesto que la grabación mostraba signos de haber sido manipulada con inteligencia artificial⁷⁶.

No obstante, sin poder ofrecer los supuestos participantes pruebas contundentes sobre su falsedad (sin duda una *probatio diabólica*⁷⁷), el audio siguió *online*. En la política de recursos manipulados de Meta se contemplaba la posibilidad de eliminar vídeos falsos en los que se mostrase a personas diciendo cosas que nunca han dicho, pero dichas directrices no incluían expresamente los audios, por lo que la grabación se encontraba en una suerte de laguna respecto a la política de la compañía.

Y no fue este el único uso de la inteligencia artificial durante este proceso electoral. Tres días antes de las elecciones, se había publicado otro audio, también del candidato liberal Michal Šimečka, en el que afirmaba que duplicaría el precio de la cerveza si ganaba las elecciones con el objetivo de luchar contra el alcoholismo. En ese caso la propia empresa Meta alertó de la existencia del audio, lo que puso a los verificadores a trabajar, consiguiendo probar en unas horas que el contenido efectivamente era falso, tal y como había declarado su protagonista⁷⁸.

Al margen de la influencia que hayan podido tener los audios sobre los resultados del proceso electoral, que perdió el candidato de Eslovaquia Progresista en contra de lo que las encuestas preveían⁷⁹, considero importante destacar tres hechos.

El primero de ellos se refiere al problema para la depuración de responsabilidades respecto de los daños creados con el uso de tecnologías *deep learning*. La tecnología utilizada hace muy difícil en primer lugar detectar que se trata de contenido falso o manipulado y, en segundo lugar, detectar el verdadero origen de los contenidos, lo que hace imposible la rendición de cuentas; de hecho, hoy en día no se conoce el

⁷⁴ Deepfakes en elecciones de Eslovaquia reafirman que IA es un peligro para la democracia. WIRED. 3 de octubre de 2023 [consulta 15 de octubre de 2024]. Disponible en <https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia>.

⁷⁵ La Agence France-Press (AFP) es una de las tres principales agencias mundiales de información, que cuenta además con un departamento de verificación de noticias falsas.

⁷⁶ *Údajná nahrávka telefonátu predsedu PS a novinárky Denníka N vykazuje podľa expertov početné známky manipulácie (La supuesta grabación de la llamada entre el presidente del PS y la periodista de Denník N muestra, según los expertos, múltiples signos de manipulación)*. AFP FACTY. 28 de septiembre de 2023 [consulta 15 de octubre de 2024]. Disponible en <https://fakty.afp.com/doc.afp.com.33WY9LF>.

⁷⁷ También llamada prueba inquisitorial, la prueba diabólica es aquella por la que se exige a la persona acusada que pruebe que algo no existe, es decir, que pruebe un hecho negativo. PASCUAL CADENA, A., *La prueba diabólica penal, entelequia normativa y prisión preventiva*. Bosch, 2021, p. 24.

⁷⁸ *Slovakia's Election Deepfakes Show AI Is a Danger to Democracy*. WIRED. 3 de octubre de 2023 [consultado el 15 de octubre de 2024]. Disponible en <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>

⁷⁹ DE NADAL, L. y JANČÁRIK, P., "Beyond the deepfake hype: AI, democracy, and "the Slovak case"", *Harvard Kennedy School (HKS) Misinformation Review*, 5(4), 2024, p. 1.

origen del audio *deep fake* sobre el candidato de EP, con el problema que ello supone en cuanto a la depuración de responsabilidades.

Esto cobra especial relevancia, y conecto ya con el segundo de los hechos a destacar, en el caso de que quien quiera manipular el proceso electoral sea una potencia extranjera. Ya en verano de 2023 la Unión Europea puso de manifiesto su preocupación por la posibilidad de injerencias extranjeras en las elecciones en el territorio de la Unión afirmando que “*la desinformación reduce la capacidad de los ciudadanos para tomar decisiones informadas y participar libremente en los procesos democráticos*” y que “*algunos Estados miembros, que Rusia considera su esfera de influencia, están más expuestos a riesgos geopolíticos derivados de la injerencia del Kremlin en su espacio de información*”⁸⁰.

De manera casi premonitoria esta resolución señalaba como posible injerencia en los procesos electorales la distribución de información distorsionada sobre los candidatos y las campañas de desinformación en redes sociales. Es decir, se sabía que podía ocurrir y, sin embargo, no pudo evitarse ni la difusión de los bulos y *deepfakes* ni sus consecuencias.

El tercero de los aspectos que debemos resaltar es el hecho de que la interlocutora del audio *deepfake* fuera una periodista, lo cual resulta importante por varios motivos. El primero de ellos es que nos encontramos ante una persona que no es un personaje público y, a pesar de ello, se ve envuelta de manera involuntaria y forzada en un asunto de corrupción electoral, con el daño que ello puede suponer para su honor y su imagen a nivel tanto personal como profesional.

El segundo motivo es que, además, esta interlocutora es una persona que se dedica profesionalmente al periodismo, resultando muy difícil pensar que esta elección haya sido azarosa, sino que más bien parece una opción premeditada que busca no solo desacreditar al candidato, quien por su dedicación a la cosa pública debe soportar un volumen mayor de injerencias sobre su honor y propia imagen, y sembrar las dudas sobre la legitimidad del proceso electoral, sino también poner en entredicho a la periodista en particular y al periodismo en su conjunto⁸¹.

Debemos en este punto recordar el lugar clave que otorga el TEDH a la prensa en los sistemas democráticos, protegiendo su papel en la formación de la opinión pública mediante la publicación de información sobre asuntos de interés público veraz y contrastada, obligando a las autoridades públicas y otras esferas de poder a rendir cuentas y señalando la corrupción y el abuso de poder⁸². Al involucrar a una periodista en una conspiración para cometer fraude electoral, se está atacando también el papel

⁸⁰ Resolución P9_TA(2023)0219 del Parlamento Europeo, de 1 de junio de 2023, sobre las injerencias extranjeras en todos los procesos democráticos de la Unión Europea, en particular la desinformación (2022/2075(INI)), Considerando T.

⁸¹ *Slovakia: Deepfake Audio Clip Aims to Manipulate Voters and Discredit Journalist Monika Tódová Ahead of Election*, Women Press Freedom, 28 de septiembre de 2023, actualizado 6 de marzo de 2024. [Consulta 15 de octubre de 2024]. Disponible en <https://www.womeninjournalism.org/threats-all/slovakia-deepfake-audio-clip-aims-to-manipulate-voters-and-discredit-journalist-monika-tdov-ahead-of-election>.

⁸² Recommendation CM/Rec (2016) 4 of the Committee of Ministers to member States on the protection of journalism and safety of journalists and other media actors.

de la prensa como vigilante público⁸³, sembrando la duda sobre su imparcialidad y su imprescindible papel en la formación de la opinión pública.

5 Conclusiones

Las herramientas de inteligencia artificial suponen una revolución tecnológica sin precedentes. Estamos ante el mayor progreso tecnológico de nuestro tiempo, equiparable o incluso de mayor relevancia que lo que supuso en su día la máquina de vapor o la electricidad⁸⁴, una revolución con efectos transversales en todas las áreas del conocimiento y de la ciencia y que nos sitúa en un nuevo escenario tecnológico, productivo y social⁸⁵.

Sin embargo, resulta extremadamente importante evaluar y prever los riesgos a los que puede someternos la IA, pues estamos ante tecnologías disruptivas que, usadas maliciosamente, tienen la potencialidad de causar daños de consecuencias catastróficas, poniendo en jaque a la democracia y hasta a la propia humanidad.

A lo largo de este trabajo se han explorado algunos de los usos de la inteligencia artificial que ya se han manifestado como amenazas reales para la libre formación de la opinión pública y, por ende, para el sistema electoral y democrático, que se asienta en el presupuesto de que la población acude a votar tras haber formado su opinión de manera libre y en base a información veraz y de calidad⁸⁶.

En el caso de la microsegmentación, nos encontramos en primer lugar con un problema en materia de protección de datos personales, dado que a menudo este tipo de prácticas adolece de falta de permisos en la recogida o tratamiento de los datos. Interesa aclarar que en el seno de la Unión Europea es un aspecto que ya se está regulando, sin que por el momento podamos sacar conclusiones sobre la efectividad de dicha regulación.

Por un lado, el Reglamento general de protección de datos⁸⁷, si bien no regula específicamente la microsegmentación, sí la limita a través de las restricciones generales en la recogida y tratamiento de los datos y en el establecimiento de ciertos principios en el tratamiento de los datos⁸⁸.

⁸³ Entre otros, asunto Pedersen and Baadsgaard c. Denmark STEDH de 17 de diciembre de 2004.

⁸⁴ ALONSO BETANZOS, A., "La inteligencia artificial y la prospectiva", *Revista Atlántica de Economía*, vol. 6, n.º 1, 2023.

⁸⁵ TUDELA ARANDA, J., "Gobierno, Parlamento, Democracia e inteligencia artificial", *Teoría y Realidad Constitucional*, 52, 2023, pp. 303-333. <https://doi.org/10.5944/trc.52.2023.39018>, p. 326.

⁸⁶ GALDÁMEZ MORALES, A., 2018, *op. cit.*, p. 31.

⁸⁷ REGLAMENTO (UE) 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 (Reglamento general de protección de datos, en adelante, RGPD).

⁸⁸ El artículo 5 RGPD establece los principios en el tratamiento de los datos que pueden resumirse en: (a) los datos personales serán tratados de manera lícita, leal y transparente en relación con el interesado («licitud, lealtad y transparencia»); (b) los datos personales solo podrán ser recogidos para propósitos especificados de antemano y no podrán usarse para ningún otro propósito («limitación de la finalidad»); (c) no deben recogerse más datos que los necesarios para la finalidad del tratamiento («minimización de datos»); (d) debe asegurarse que los datos sean exactos («exactitud»); (e) los datos personales no serán almacenados más tiempo del necesario («limitación del plazo de conservación»); (f) los datos personales deben estar protegidos («integridad y confidencialidad»).

Sí regula de manera específica la segmentación en materia electoral el reciente Reglamento sobre transparencia y segmentación en publicidad política⁸⁹, cuyo considerando 6 advierte de que la microsegmentación y otras técnicas avanzadas de tratamiento de los datos “pueden plantear graves amenazas para intereses públicos legítimos como la equidad, la igualdad de oportunidades y la transparencia en los procesos electorales y los derechos fundamentales de libertad de expresión, privacidad y protección de los datos personales, así como de igualdad y no discriminación y el derecho fundamental a ser informado de manera objetiva, transparente y plural”.

Este reglamento permite la segmentación siempre que se cumplan tres requisitos: 1) que el interesado haya proporcionado los datos personales al responsable del tratamiento; 2) que el interesado haya dado, por separado, su consentimiento expreso al tratamiento de los datos personales a efectos de publicidad política, y 3) que dichas técnicas no impliquen la «elaboración de perfiles»⁹⁰.

Por último, el Reglamento de inteligencia artificial⁹¹ prohíbe expresamente en su artículo 5.1.a) bajo pena de sanción de hasta 35 millones de euros (a concretar por los Estados miembros) “la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que adopten una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas”, lo que parece que deja fuera de la legalidad el marketing cognitivo.

Además, se califican como de “alto riesgo” los sistemas de IA que elaboren perfiles de personas físicas y los sistemas de IA destinados a ser utilizados para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos.

Sin duda, toda esta reciente regulación está orientada a limitar los riesgos de los usos ilícitos de las herramientas de inteligencia artificial y es mucho más de lo que se ha hecho en otros entornos legislativos, pero habrá que ver si es suficiente o si los intereses en los resultados que pueden obtenerse infringiendo la normativa compensan el riesgo.

Por otro lado, los sistemas y herramientas de inteligencia artificial pueden ser utilizados para llevar a cabo campañas de desinformación masiva que, aunque no estén centradas de manera directa en el ámbito electoral, sí afectarán a la formación de la opinión pública. Ya sea partiendo de contenidos fabricados por humanos o

⁸⁹ REGLAMENTO (UE) 2024/900 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 13 de marzo de 2024 sobre transparencia y segmentación en la publicidad política.

⁹⁰ Artículo 18 del Reglamento sobre transparencia y segmentación en la publicidad política.

⁹¹ REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los reglamentos (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de inteligencia artificial).

fabricados por máquinas, la inteligencia artificial facilita la difusión absolutamente viral del contenido apoyada por cuentas automatizadas que pueden incluso contestar comentarios de manera indistinguible a como lo haría un humano y adaptando el contenido del mensaje en la medida necesaria para causar efecto en la persona destinataria.

Esta facilidad en la difusión se ve favorecida por un escenario en el que los medios de comunicación han perdido su papel como intermediarios de la información, siendo sustituidos por las redes sociales, donde circula todo tipo de información publicada por infinidad de fuentes, también por los propios medios de comunicación, lo que dificulta al ciudadano de a pie separar la información de calidad de la mera desinformación.

A ello se suma el hecho de que no se trata solo de que el número de informadores haya crecido exponencialmente, mezclándose medios de comunicación con *pseudo-medios*, *influencers* y opinadores en general, ni siquiera de que la labor de contrastación sea mejor, peor o directamente inexistente, es que, además, las herramientas de *deep learning* permiten fabricar las pruebas audiovisuales que se necesitan para apoyar la información que se quiere difundir, como ocurrió en el caso eslovaco.

La solución que ofrece el Reglamento de inteligencia artificial de la Unión Europea para este asunto pasa por aumentar la transparencia, obligando a los proveedores que generen contenido sintético de audio, imagen, vídeo o texto a marcar la información de salida del sistema de IA en un formato legible por máquina para que sea posible detectar que ha sido generada o manipulada de manera artificial⁹².

Es decir, se trata de incorporar como metadatos o mediante datos cifrados en una capa invisible al ojo humano la información de que se trata de contenido generado por una IA. Pero ¿esto va a impedir que el contenido se replique como veraz por cientos o miles de personas? Ciertamente no, pero será posible probar que es falso, al menos hasta que alguien con los conocimientos necesarios elimine la marca de agua.

No obstante todo lo anterior, resulta importante tener en cuenta que, con todo lo disruptiva que es la inteligencia artificial en el momento actual, se encuentra en los albores de su existencia; lo que nos espera supera con creces lo que hemos visto hasta el momento. Así, desde la doctrina científica se han señalado tres grandes etapas o tipos de inteligencia artificial que representan distintos niveles de autonomía y habilidad en las máquinas⁹³:

Inteligencia artificial estrecha (ANI, por sus siglas en inglés): se refiere a sistemas de IA diseñados y entrenados para realizar tareas específicas y limitadas. Estas inteligencias no poseen conciencia, comprensión o verdaderas capacidades cognitivas humanas. Operan dentro de un conjunto restringido de

⁹² Artículo 50, apartado 2 del REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los reglamentos (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de inteligencia artificial).

⁹³ SERRANO VALVERDE, Á., "Tecnología que preocupa: de la IA a Gutenberg", *EME Experimental Illustration, Art & Design*, (12), 2024, pp. 64–75. <https://doi.org/10.4995/eme.2024.21212>, p. 65.

parámetros y no pueden generalizar su aprendizaje a dominios fuera de su programación original.

Inteligencia artificial general (AGI, por sus siglas en inglés): describe una IA capaz de entender, aprender y aplicar su inteligencia para resolver cualquier problema de forma similar a como lo haría un ser humano. Una AGI tendría habilidades cognitivas generales, permitiéndole transferir conocimientos y habilidades entre diferentes dominios y tareas. Hasta la fecha, la mayoría de la literatura científica coincide en que la AGI no ha sido lograda, aunque con cierto debate⁹⁴.

Inteligencia artificial superinteligente (ASI, por sus siglas en inglés): se refiere a una hipotética IA que supera la inteligencia humana en todos los aspectos: creatividad, sabiduría, resolución de problemas y habilidades sociales, entre otros. La ASI no solo podría realizar todas las tareas intelectuales que un ser humano puede hacer, sino que lo haría de manera más eficiente y efectiva. Este nivel de inteligencia artificial plantea importantes consideraciones éticas y de seguridad, ya que su comportamiento y decisiones podrían estar más allá de la comprensión humana.

Sin duda, las posibilidades son prácticamente infinitas y el desarrollo de esta tecnología tiene el potencial de resolver los grandes problemas de la humanidad, desde el cambio climático a la escasez de recursos. Cuando se desarrolle la superinteligencia artificial estaremos en condiciones de resolver cualquier problema que se nos plantee; la cuestión es que una tecnología así de potente puede resultar peligrosa incluso aunque se la utilice para un objetivo inocuo o incluso beneficioso *a priori* para la humanidad y el planeta.

Sobre ello teorizó el filósofo Nick Bostrom, empleando la hipótesis de una IA cuyo objetivo sea maximizar la producción de clips. En el desarrollo de su tarea, esta IA iría mejorando los procesos de producción, pero podría suceder que, buscando fabricar cada vez más y más clips, comenzase a transformar toda la materia de la Tierra y después porciones cada vez más grandes del espacio en plantas de producción de clips⁹⁵.

De igual manera, utilizar un sistema de IA para conseguir un objetivo en principio legítimo como, por ejemplo, diseñar o mejorar una campaña electoral o mejorar la opinión de un candidato/a en las encuestas, podría fácilmente derivar en una máquina que, en busca de cumplir su objetivo, no dudara en utilizar todos los medios a su disposición para lograrlo.

Esto podría implicar no solo la elaboración de perfiles y diseño de propaganda electoral ultrapersonalizada, sino también la fabricación de cualquier contenido que

⁹⁴ A este respecto, resulta interesante el análisis de Sébastien Bubeck y otros sobre la nueva versión de la aplicación Chat GPT, que les lleva a afirmar que GPT-4 muestra destellos de inteligencia artificial general en sus capacidades mentales fundamentales (como el razonamiento, la creatividad y la deducción), la variedad de temas en los que ha adquirido experiencia (como la literatura, la medicina y la programación) y la diversidad de tareas que es capaz de realizar (por ejemplo, jugar, usar herramientas, explicarse a sí mismo...). BUBECK, S. y otros, *Sparks of Artificial General Intelligence. Early experiments with GPT-4*. Arxiv, 2023, <https://doi.org/10.48550/arXiv.2303.12712>.

⁹⁵ BOSTROM, N., *Ethical Issues in Advanced Artificial Intelligence*, 2003, Disponible en <https://nickbostrom.com/ethics/ai> (consulta 24 de octubre de 2024).

fuera necesario para apoyar el argumentario o destruir a los oponentes, la desinformación sobre el propio proceso electoral para lograr la abstención involuntaria de ciertos sectores de la población o, incluso, la modificación de antecedentes penales, expedientes académicos o historiales médicos, todo cuanto fuera necesario para lograr su objetivo⁹⁶.

El principal abordaje a este problema pasa actualmente por la alineación (*alignment*, en inglés), tratar de alinear los intereses y objetivos de la IA con los intereses y objetivos de sus diseñadores en particular y con los de los seres humanos en general⁹⁷ y establecer los guardarrailes necesarios para evitar que el sistema de IA tome decisiones peligrosas o indeseables.

El problema es que, a medida que los sistemas de IA se hacen más potentes y la tecnología va evolucionando hacia la inteligencia artificial general y la superinteligencia artificial, más difícil resultará establecer de manera efectiva ese alineamiento, puesto que la máquina estará en condiciones de prever e inventar muchos más escenarios de los que puedan prever los seres humanos que la programen.

Se hace evidente que las herramientas con las que contamos no son suficientes para enfrentarnos a los retos que vienen. En un escenario político convulso como el que vivimos, con conflictos armados abiertos en múltiples puntos del planeta y con intereses políticos radicalmente enfrentados, la desinformación y, más concretamente, la desinformación haciendo uso de herramientas de inteligencia artificial, se configura como una clara amenaza para la libre formación de la opinión pública y, con ello, para el ejercicio libre del derecho de sufragio⁹⁸ y el procedimiento electoral en su conjunto, pudiendo suponer un golpe en la línea de flotación de los sistemas democráticos tal y como los conocemos.

6 Bibliografía

ABA CATOIRA, A., "Libertades de expresión e información en la sociedad digital como garantías de la democracia", *Revista Doctrina Distrital*, vol. 1, n.º extra 2, 2021, pp. 55-70.

ABA CATOIRA, A., "Los desórdenes informativos en un sistema de comunicación democrático", *Revista de Derecho Político*, 1(109), 2020, pp. 119151.

<https://doi.org/10.5944/rdp.109.2020.29056>.

⁹⁶ Más sobre la vulnerabilidad del mundo ante los avances de la IA en BOSTROM, N., "The Vulnerable World Hypothesis" *Global Policy*, vol. 10:4, 2019, pp. 455-476.

⁹⁷ Este problema fue formulado en 1960 por Norbert Wiener del siguiente modo: "Si utilizamos, para alcanzar nuestros objetivos, un medio mecánico en cuyo funcionamiento no podemos interferir de manera efectiva una vez que lo hemos iniciado, porque la acción es tan rápida e irrevocable que no tenemos los datos para intervenir antes de que la acción esté completa, entonces es mejor que estemos completamente seguros de que el propósito que hemos introducido en la máquina es el propósito que realmente deseamos y no simplemente una imitación colorida del mismo". WIENER, N., "Some Moral and Technical Consequences of Automation", *Science*, 131(3410), 1960, pp. 1355-1358.
<http://www.jstor.org/stable/1705998>. DOI: 10.1126/science.131.3410.1355 p. 1358.

⁹⁸ PAUNER CHULVI, P., 2018, *op. cit.*, p. 299: "La desinformación afecta directamente a los fundamentos de las sociedades democráticas que se construyen sobre la presunción de que sus ciudadanos toman decisiones informadas en el momento de acudir a votar".

- ALONSO BETANZOS, A., “La inteligencia artificial y la prospectiva”, *Revista Atlántica de Economía*, vol. 6, n.º 1, 2023.
- ANDERSON, B. y HORVARTH, B., PÉREZ LAGÜELA, E. (trad.). “El ascenso de la máquina de armas propagandísticas de inteligencia artificial”, *Papeles de relaciones ecosociales y cambio global*, 38, 2017, pp. 37-47.
- ASTUDILLO MUÑOZ, J. L., “Notas sobre la posverdad, sus efectos en el sistema democrático y en la protección de los derechos humanos en el marco de la sociedad digital”, *Teoría y Realidad Constitucional*, 52, 2023, pp. 401-427.
<https://doi.org/10.5944/trc.52.2023.39023>
- BARRIENTOS-BÁEZ, A., CALDEVILLA-DOMÍNGUEZ, D. y YEZERS'KA, L., “Fake news y posverdad: Relación con las redes sociales y fiabilidad de contenidos”, *Fonseca, Journal of Communication*, (24) 2022, pp. 149-162.
<https://doi.org/10.14201/fjc.28294>
- BELL, E., “Facebook se está comiendo el mundo”, *Cuadernos de Periodistas*, 32 2016, pp. 67-76. <https://www.cuadernosdeperiodistas.com/facebook-se-esta-comiendo-mundo/>
- BERROCAL-GONZALO, S., WAISBORD, S. y GÓMEZ-GARCÍA, S., “Polarización política y medios de comunicación, su impacto en la democracia y la sociedad”, *Profesional de la información*, v. 32, n. 6, e320622, 2023, <https://doi.org/10.3145/epi.2023.nov.22>
- BOSTROM, N., *Ethical Issues in Advanced Artificial Intelligence*, 2003, Disponible en <https://nickbostrom.com/ethics/ai> (consulta 24 de octubre de 2024).
- BOSTROM, N., “The Vulnerable World Hypothesis”, *Global Policy*, vol. 10:4, 2019, pp. 455-476.
- BUBECK, S. y otros, *Sparks of Artificial General Intelligence. Early experiments with GPT-4*. Arxiv, 2023, <https://doi.org/10.48550/arXiv.2303.12712>
- CAMARGO ORTEGA, J. F., CAMARGO VEGA, J. J., & JOYANES AGUILAR, L., “Conociendo Big Data”, *Revista FI-UPTC*, 24(38), 2015, pp. 63-77.
- CARMONA SALGADO, C., *Libertad de expresión e información y sus límites*, EDERSA Editoriales de Derecho Reunidas, 1991.
- DE NADAL, L. y JANČÁRIK, P., “Beyond the deepfake hype: AI, democracy, and “the Slovak case””, *Harvard Kennedy School (HKS) Misinformation Review*, 5(4), 2024. *Deepfakes en elecciones de Eslovaquia reafirman que IA es un peligro para la democracia*. WIRED. 3 de octubre de 2023 [consulta 15 de octubre de 2024]. Disponible en <https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia>
- DEL-FRESNO-GARCÍA, M., “Desórdenes informativos: sobreexuestos e infrainformados en la era de la posverdad”, *El profesional de la información*, v. 28, n. 3, 2019, e280302. <https://doi.org/10.3145/epi.2019.may.02>.
Digital Report 2024 disponible en <https://datareportal.com/reports/digital-2024-global-overview-report> [consulta 17 de octubre de 2024]
- ELVIRA PERALES, A., “¿Quién es periodista?”, *Teoría y Realidad Constitucional*, 52, 2023, pp. 209-231. <https://doi.org/10.5944/trc.52.2023.39015>

Regap



ESTUDIOS

- ENISA, EEAS. *Foreign Information Manipulation Interference (FIMI) and Cybersecurity - Threat Landscape*. DOI: 10.2824/750. (2022) Disponible en https://www.cde.ual.es/wp-content/uploads/2022/12/EEAS-ENISA-Disinformation_Misinformation.pdf
- FESTINGER; L., “Teoría de la disonancia cognoscitiva (Extractado por A OVEJERO)”, *Psicothema*, 5(1), 1993, pp. 201-206. Recuperado a partir de <https://reunido.uniovi.es/index.php/PST/article/view/7156>
- GALDÁMEZ MORALES, A., “Posverdad y crisis de legitimidad: El creciente impacto de las fake news”, *Revista Española de la Transparencia*, vol. 8, 2018, pp. 25-44. <https://doi.org/10.51915/ret.45>
- GARCÍA OROSA, B., GALLUR SANTORUN, S. y LÓPEZ GARCÍA, X., “El uso del clic-bait en cibermedios de los 28 países de la Unión Europea”, *Revista Latina de Comunicación Social*, 72, 2017, pp. 1.261-1.277. <http://www.revistalatinacs.org/072paper/1218/68es.html>
- GARCÍA MARÍN, D., “Infodemia global. Desórdenes informativos, narrativas fake y fact-checking en la crisis de la Covid-19”, *Profesional de la información*, v. 29, n. 4, 2020, e290411. <https://doi.org/10.3145/epi.2020.jul.11>
- GARRIGUES WALKER, A. y GONZÁLEZ DE LA GARZA, L. M., *El Derecho a no ser engañado. Y cómo nos engañan y nos autoengañamos*. Aranzadi, 2020.
- GONZÁLEZ DE LA GARZA, L. M., “Por qué las campañas electorales cognitivas basadas en la posverdad pueden erosionar las democracias de opinión”, *Revista General de Derecho Constitucional*, 32, 2020.
- GUAÑA-MOYA, J., SALGADO-REYES, N. y ESCOBAR-TERÁN, C., “La informática y la ciencia de datos utilizando Deep learning. [Computing and data science using deep learning]”, *Revista Ibérica De Sistemas e Tecnologías De Informação*, 2022, pp. 116-127. <https://www.proquest.com/scholarly-journals/la-informática-y-ciencia-de-datos-utilizando-deep/docview/2714752347/se-2>
- HAN, B., *Infocracia. La digitalización y la crisis de la democracia*. Ed. Taurus. (2022).
- ISAAK, J. y HANNA, M. J., “User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection”, in *Computer*, vol. 51, no. 8, 2018, pp. 56-59, August 2018, doi: 10.1109/MC.2018.3191268.
- JIMÉNEZ PALOMARES, P. y GOMES-FRANCO E SILVA, F., “Visibilidad de la información en redes sociales: los algoritmos de Facebook y su influencia en elclickbait”, *Caleidoscopio - Revista Semestral de Ciencias Sociales y Humanidades*, 23(41), 2019, pp. 173-211. doi:10.33064/41crscsh1772C
- LEÓN-MENDOZA, R., “La imagen como forma de (des)conocimiento en la era del deepfake”, *ANIAV - Revista de Investigación en Artes Visuales*, n. 11, 2022, pp. 53-70, septiembre. <https://doi.org/10.4995/aniav.2022.17309>
- LEVI, S., *#FakeYou: Fake news y desinformación*, Rayo Verde Editorial, 2020.
- LINARES RODRÍGUEZ, V. y ABEJÓN MENDOZA, P., “De la marca Obama 2.0 al antagonismo digital de Donald Trump”, *Investigación y desarrollo*, 29(1), 2021, pp. 99-132. <https://doi.org/10.14482/indes.29.1.658.8>

- MARTÍN-HERRERA, I. y GUERRERO-CABALLERO, M., “Millennials y Generación X frente a la realidad del Big data y la protección de datos personales en Internet”, *Vivat Academia*, 157, 2024, pp. 1-20.
<http://doi.org/10.15178/va.2024.157.e1489>
- MARTÍN, M., FIOR, A., COSTA ESCUREDO, Á., RUIZ SAN ROMÁN, J. A., “La figura del periodista frente al fenómeno de la desinformación en la era digital”, *Zer*, 29(56), 2024, pp. 65-85. (<https://doi.org/10.1387/zer.25993>).
- MÉNDEZ, L., “El algoritmo sustituye al periodismo”, *Cuadernos de Periodistas*, 2021, pp. 65-70, (2018). Disponible en <https://www.cuadernosdeperiodistas.com/el-algoritmo-sustituye-al-periodismo/>
- MENEU-BORJA, M., SORIA SALVADOR, A., AGUAR TORRES, J., “La juventud en el ecosistema mediático: una mirada al consumo y la producción informativos de las nuevas generaciones”, *Zer*, 29(56), 2024, pp. 153-172.
(<https://doi.org/10.1387/zer.26192>).
- MORENO-CABANILLAS, A., CASTILLERO-OSTIO, E. y SERNA-ORTEGA, Á., “El impacto de las redes sociales en la campaña política: Elecciones Generales de 2023 en España”, *Redmarka. Revista de Marketing Aplicado*, vol. 28, núm. 1, 2024, pp. 56-76. <https://doi.org/10.17979/redma.2024.28.1.10>
- Oficina de Ciencia y Tecnología del Congreso de los Diputados (Oficina C). Informe C: Desinformación en la era digital. www.doi.org/10.57952/j3p6-9086 (2023).
- PALAU-SAMPIO, D., CARRATALÁ, A., “Injecting disinformation into public space: pseudo-media and reality-altering narratives”, *Profesional de la información*, v. 31, n. 3, 2022, e310312. <https://doi.org/10.3145/epi.2022.may.12>.
- PASCUAL CADENA, A., *La prueba diabólica penal, entelequia normativa y prisión preventiva*, Bosch, 2021.
- PASTOR-GALINDO, J. *et al.*, “Twitter social bots: The 2019 Spanish general election data”, *Data in Brief*, Volume 32, 2020.
<https://doi.org/10.1016/j.dib.2020.106047>
- PASTOR-GALINDO, J. *et al.*, “Spotting Political Social Bots in Twitter: A Use Case of the 2019 Spanish General Election”, *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, 2020, pp. 2156-2170, doi: 10.1109/TNSM.2020.3031573.
- PAUNER CHULVI, C., “Noticias falsas y libertad de expresión e información. El control de los contenidos informativos en la red”, *Teoría y Realidad Constitucional*, 41, pp. 297-318. <https://doi.org/10.5944/trc.41.2018.22123>
- RODRÍGUEZ PÉREZ, C., “No diga fake news, di desinformación: una revisión sobre el fenómeno de las noticias falsas y sus implicaciones”, *Revista Comunicación*, 40, 2019, pp. 65-74, DOI: <http://dx.doi.org/10.18566/comunica.n40.a05>
- RUBIO NÚÑEZ, R., “Quiero ser como Obama (me pido una red social)”, *Cuadernos de pensamiento político* (Madrid, Spain), 21, 2009, pp.123-154.
- SÁDABA RODRÍGUEZ, I. y GARCÍA ARNAU, A., “Teorías de la conspiración y crisis globales: retos para la teoría sociológica”, *RES. Revista Española de Sociología*, 33(2), 2024. <https://doi.org/10.22325/fes/res.2024.222>

Regap



ESTUDIOS

- SANTOS SANZ, L. “Identificación automática de bots en Twitter basada en contexto”, *Reports de inteligencia económica y relaciones internacionales*, n.º 13 (Vol.), 2023, <http://hdl.handle.net/10486/700265>
- SERRANO VALVERDE, Á., “Tecnología que preocupa: de la IA a Gutenberg”, *EME Experimental Illustration, Art & Design*, (12), 2024, pp. 64-75. <https://doi.org/10.4995/eme.2024.21212>
- Slovakia: Deepfake Audio Clip Aims to Manipulate Voters and Discredit Journalist Monika Tódová Ahead of Election, *Women Press Freedom*, 28 de septiembre de 2023, actualizado 6 de marzo de 2024. [Consulta 15 de octubre de 2024]. Disponible en <https://www.womeninjournalism.org/threats-all/slovakia-deepfake-audio-clip-aims-to-manipulate-voters-and-discredit-journalist-monika-tdov-ahead-of-election>
- Slovakia’s Election Deepfakes Show AI Is a Danger to Democracy. WIRED. 3 de octubre de 2023 [consultado el 15 de octubre de 2024]. Disponible en <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>
- SUPASORN SUWAJANAKORN, STEVEN M. SEITZ, AND IRA KEMELMACHER-SHLIZERMAN. *Synthesizing Obama: Learning Lip Sync from Audio*. ACM Trans. Graph. 36, 4, Article 95 (July 2017), 13 pages. (2017) DOI: <http://dx.doi.org/10.1145/3072959.3073640>
- TEJEROMATOS, P., “¿Populismo híbrido? Análisis del SMER-SD en Eslovaquia”, *Revista Jurídica de La Universidad Autónoma de Madrid*, 44, 2021, pp. 175-198. <https://doi.org/10.15366/rjuam2021.44.006>
- TERNOVSKI, J., KALLA, J. y ARONOW, P. M., *Deepfake Warnings for Political Videos Increase Disbelief but Do Not Improve Discernment: Evidence from Two Experiments*. <https://doi.org/10.31219/osf.io/dta97> (2021).
- TORRES DEL MORAL, A. (dir.), *Libertades informativas*, Colex, 2009.
- TUDELA ARANDA, J., “Gobierno, Parlamento, Democracia e inteligencia artificial”, *Teoría y Realidad Constitucional*, 52, 2023, pp. 303-333. <https://doi.org/10.5944/trc.52.2023.39018>
- Údajná nahrávka telefonátu predsedu PS a novinárky Denníka N vykazuje podľa expertov početné známky manipulácie (La supuesta grabación de la llamada entre el presidente del PS y la periodista de Denník N muestra, según los expertos, múltiples signos de manipulación). AFP FACTY. 28 de septiembre de 2023 [consulta 15 de octubre de 2024] Disponible en <https://fakty.afp.com/doc.afp.com.33WY9LF>
- WIENER, N., “Some Moral and Technical Consequences of Automation”. *Science*, 131(3410), 1960, pp. 1355-1358. DOI: [10.1126/science.131.3410.1355](https://doi.org/10.1126/science.131.3410.1355)
- WILLOCQ, S., “Explaining Time of Vote Decision: The Socio-Structural, Attitudinal, and Contextual Determinants of Late Deciding”, *Political Studies Review*, 17(1), 2019, pp. 53-64. <https://doi.org/10.1177/1478929917748484>
- WOOLLEY, S. C. y HOWARD, P. N., *Computational Propaganda Worldwide: Executive Summary*. Working Paper 2017.11. Oxford, UK: Project on Computational Propaganda. demtech.oii.ox.ac.uk. 14 pp.