



Revista Galega de Administración Pública, EGAP
Núm. 68_xullo-decembro 2024 | pp. 89-114
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 predoutoral de Dereito Constitucional
Universidade da 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 A importancia da libre formación da opinión pública. 3 A manipulación do electorado a través de propaganda cognitiva e *microtargeting*: o caso de Cambridge Analytica. 4 O uso da intelixencia artificial para atacar os candidatos: o caso de Eslovaquia. 5 Conclusións. 6 Bibliografía.

1 Introducción

Vivimos nun momento complexo para a información. Pasamos dun escenario en que tan só uns poucos controlaban a elaboración de noticias e as puñan á disposición do público, primeiro a través da prensa impresa e despois a través da radio e a televisión, a outro moito máis dinámico, extenso e con doado acceso para calquera persoa¹.

Este doado acceso maniféstase tanto desde o punto de vista activo como pasivo. Así, desde o punto de vista activo, atopámonos con que a universalización no uso de Internet e o desenvolvemento das redes sociais cambiou o propio concepto de informador, ao non resultar xa necesario pasar polo filtro dunha dirección ou unha liña editorial para poder publicar, nin sequera cursar unha formación ou especialización habilitante, senón que calquera persoa ten a posibilidade de difundir calquera

¹ Neste sentido, sinala o *Informe C: Desinformación en la era digital* da Oficina de Ciencia e Tecnoloxía do Congreso que, aínda que Internet e o desenvolvemento dixital propician grandes avances e beneficios económicos e sociais, tamén deron pé a “un novo contexto social e informativo que favoreceu unha amplificación sen precedentes da desinformación e dos seus efectos, converténdoa nunha destacada ameaza para os sistemas democráticos. Trátase dunha cuestión de seguridade nacional que alcanza niveis críticos en situacións de gran relevancia social, como crises sanitarias, conflitos bélicos ou os procesos electorais”. Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C). Informe C: Desinformación na era dixital. www.doi.org/10.57952/j3p6-9086 (2023) p. 1/26.

contido e comentar ou replicar a información publicada chegando a unha audiencia antes impensable fóra dos medios tradicionais².

Isto supón unha multiplicación e diversificación do informador debido á socialización tecnolóxica da produción de contidos e, polo tanto, á democratización da transmisión da información³ que fai que cada vez sexa máis difícil distinguir entre o xornalismo profesional e a información partidista, e entre medios de comunicación profesionais e medios alternativos⁴ ou *pseudomedios*⁵.

Desde o punto de vista pasivo, atopámonos cun consumidor de información voraz, que ten un océano de contidos ao alcance da súa man⁶ ao que con grande acerto se refire Carlos Rodríguez como *catering* de fontes de información⁷, e que mostra cada vez maior tendencia ao consumo compulsivo de información, outorgándolle en moitas ocasións máis importancia á rapidez e á accesibilidade que á calidade e á fiabilidade⁸.

Aquí entra en xogo a lei da oferta e a demanda, provocando que, nun ecosistema en que o consumidor quere consultar información sobre un feito ao instante de que este aconteza, algo claramente incompatible co dilixente labor xornalístico de busca de información e contrastación da noticia, moitos medios non dubidasen en pregarse á vontade dos seus lectores, chegando a arriscar a veracidade e o rigor das súas publicacións a cambio de conseguir a primicia, a exclusiva, a cambio de ser os primeiros en ofrecerlle ao público unha información⁹.

Deste xeito, o deber de contrastación da noticia, ese que tradicionalmente evitara a publicación de informacións falsas ou mesmo salvara a responsabilidade de medios e xornalistas pola publicación de información errónea, vai quedando pouco a pouco

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., e 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., e 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: sobreexposados e infrainformados 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 Considéranse como *pseudomedios* os sitios web que imitan o estilo e deseño utilizado polos medios e xornalistas *mainstream*, pero con fins antagónicos á ética xornalística, mesturando datos e opinión, todo iso cunha gran carga ideolóxica. Trátase de webs que intentan mostrarse como unha alternativa aos medios convencionais pero que moitas veces se limitan a absorber o contido destes e reformulalo de acordo cunha ideoloxía concreta: PALAU-SAMPIO, 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 Segundo o Digital Report 2024, o número de usuarios de teléfonos móbiles ascendía a principios de 2024 a 5,61 billóns, un 69,4% da poboación mundial, sufrindo un incremento de 138 millóns desde comezos de 2023. Informe dispoñible en <https://datareportal.com/reports/digital-2024-global-overview-report> [consulta 17 de outubro 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. Dispoñible en <https://www.cuadernosdeperiodistas.com/el-algoritmo-sustituye-al-periodismo/>

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

relegado a un segundo ou terceiro plano, eclipsado pola busca de *likes*, *retuits* e de páxinas vistas¹⁰.

É que, malia ser certo que a nosa Constitución no seu artigo 20.1.d) impón o límite da veracidade á información ao consagrar o dereito a comunicar ou recibir libremente información veraz por calquera medio de difusión, non se trata dunha exigencia absoluta, que suporía case unha forma de censura¹¹, senón que o que se exige é un deber de dilixencia na comprobación dos feitos, evitando que o comunicador actúe con menosprezo á veracidade ou falsidade do comunicado, comportándose de forma negligente e irresponsable ao transmitir como feitos verdadeiros simples rumores carentes de toda constatación ou meras invencións, rumores ou insinuacións¹².

Non obstante, no contexto actual, atopámonos con que non só non se prima ou se valora a veracidade, senón que mesmo se está a intentar anular o propio concepto de verdade, converténdoa en algo subxectivo. É o que se veu a chamar *posverdade*¹³, definida pola Real Academia Española como a “distorsión deliberada dunha realidade, que manipula crenzas e emocións co fin de influír na opinión pública e en actitudes sociais”.

Todo pode ser verdade ou non segundo o relato que cada un queira crer, quedando a verdade relativizada e subordinada ás emocións e á realidade desexada e totalmente desvinculada da realidade fáctica¹⁴. E iso porque, se nada é verdade, dificilmente pode ser algo mentira, quedando o campo fertilizado para que se poida difundir a información que máis interese en cada momento para condicionar as decisións da poboación en xeral na dirección desexada.

Adentrámonos, polo tanto, nun contexto de *infocracia* ou réxime da información¹⁵ onde priman a abundancia de contidos sobre a súa veracidade, un mal que

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 a Sentenza do Tribunal Constitucional 6/1988, do 21 xaneiro, que «as afirmacións erróneas son inevitables nun debate libre, de tal forma que, de se impor “a verdade”, como condición para o recoñecemento do dereito, a única garantía da seguridade xurídica sería o silencio».

¹² Sobre o requisito da veracidade, sentenzas do Tribunal Constitucional 240/1992, do 21 de decembro, FX 5.º e 7.º, e 105/1990, do 6 de xuño, FX. 5º.

¹³ O termo *posverdade* foi popularizado polo expresidente estadounidense Donald Trump en 2016, chegando a ser acuñado como palabra dese ano por *Oxford Dictionaries*. Non obstante, a primeira aparición da palabra *posverdade* débemoslla ao guionista e dramaturgo Steve Tesich, que a utilizou en 1992 nun artigo para a revista *The Nation* sobre o escándalo Watergate e a guerra de Iraq, afirmando «Lamento que nós, como pobo libre, decidísemos libremente vivir nun mundo onde reina a posverdade».

¹⁴ Afirmar Jorge Astudillo que “o que caracteriza a posverdade é que os feitos obxectivos teñen menos credibilidade ou menos influencia que os sentimentos, desexos e crenzas das persoas ao momento de asumir unha posición ou emitir unha opinión sobre un tema de interese público ou dar o seu apoio ou rexeitamento a unha 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. e YEZERS'KA, L., 2022, *op. cit.*, 153.

¹⁵ Sobre o termo *infocracia*, sinala Byung-Chul Han que a democracia está a dexenerar en *infocracia* debido a que estamos sendo arrastrados por un tsunami de información da esfera política que provoca distorsións e trastornos masivos no proceso democrático debido en gran parte á dixitalización do mundo en que vivimos, que avanza inexorable sometendo a nosa percepción, a nosa relación co mundo e a nosa convivencia a un cambio radical. Define este autor o réxime da información como “a forma de dominio en que a información e o seu procesamento mediante algoritmos e intelixencia artificial determinan de modo decisivo os procesos sociais, económicos e políticos HAN, B., *Infocracia. La digitalización y la crisis de la democracia*, Ed. Taurus, 2022, p. 25. e p. 10.

se fixo especialmente patente durante a pandemia da covid-19, en que a propia Organización Mundial da Saúde afirmou que a súa loita se libraba tanto contra a pandemia como contra a *infodemia*¹⁶ para referirse á abundancia tal de información sobre un asunto concreto que resulta difícil discriminar a información verdadeira da mera desinformación¹⁷.

Neste escenario, propicio de seu para que campe a desinformación con total liberdade, irrompe a intelixencia artificial¹⁸ para revolucionalo todo, a introducir supostos que ata agora eran impensables e que supoñen un reto ao labor de contrastación dos profesionais do xornalismo, pero, sobre todo e máis importante, supón un gran risco para a cidadanía de a pé, para quen distinguir contidos veraces de contidos artificiais se torna máis difícil que nunca, por non dicir imposible.

Isto cobra especial relevancia no marco do proceso electoral, un proceso que debe contar con todas as garantías para que a participación da cidadanía teña lugar de xeito libre e informado e a través de sufraxio libre, igual, directo e secreto e onde os medios de comunicación representaban un importante papel, pondo ao dispor do electorado información plural e veraz para unha libre formación da opinión pública.

A importancia dos efectos da desinformación sobre os procesos electorais deriva precisamente da transcendencia que ten o proceso electoral, mediante o cal a cidadanía elixe as persoas que a van representar nas distintas instancias durante os próximos catro anos, sen capacidade de volver atrás esa decisión nin de revogar o poder de representación outorgado, mesmo no caso de que na formación da súa decisión en favor dunha ou outra opción interviñese engano ou mentira en forma de desinformación.

Ademais, no momento actual, en que a fragmentación das opcións electorais e a propia desafección da cidadanía cara á política pode provocar un atraso na decisión do voto¹⁹, a información ou desinformación publicada nos días previos á votación resulta máis relevante, inclinando definitivamente a balanza do electorado indeciso en favor dunha ou outra opción.

Nun contexto en que as desordes informativas son a norma, realmente somos libres nas nosas decisións? Podemos seguir falando dunha libre formación da opinión

¹⁶ Así o sinalou o director xeral da OMS na Conferencia de Seguridade de Múnic o 15 de febreiro de 2020: "Pero non estamos a loitar unicamente contra unha epidemia; estamos a loitar contra unha infodemia. As noticias falsas propáganse con máis rapidez e facilidade que o propio virus, e son igual de perigosas". Dispoñible en <https://www.who.int/es/director-general/speeches/detail/munich-security-conference>.

¹⁷ Sinala David García-Marín que, aínda que o termo foi popularizado pola OMS, "xa en 2003 este termo se utilizou para describir situacións onde un conxunto de feitos, mesturados con medo, especulación e rumores, son amplificados e distribuídos a unha audiencia mundial grazas ao uso das tecnoloxías da información, con posibilidades de impactar de forma desproporcionada na seguridade, economía e política dos 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 O Regulamento europeo de intelixencia artificial define un sistema de IA como un sistema baseado nunha máquina que está deseñado para funcionar con distintos niveis de autonomía e que pode mostrar capacidade de adaptación tras o despregamento, e que, para obxectivos explícitos ou implícitos, infire da información de entrada que recibe o xeito de xerar resultados de saída, como predicións, contidos, recomendacións ou decisións, que poden influír en ámbitos físicos ou virtuais.

¹⁹ 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>, p. 59.

pública? Podería unha potencia estranxeira influír na elección dos nosos máximos representantes? Que papel pode desempeñar a intelixencia artificial en todo isto?

2 A importancia da libre formación da opinión pública

A libre formación da opinión pública²⁰ considérase como un elemento intrínseco ao sistema democrático, unha institución indisolublemente ligada co pluralismo político, consagrado polo artigo 1 da nosa Constitución como valor fundamental e requisito do funcionamento do Estado de dereito, xa que non hai opinión pública sen democracia, nin democracia sen opinión pública²¹.

Considera o noso Tribunal Constitucional que as liberdades do artigo 20 da Constitución aparecen como unha garantía da opinión pública libre, sen a cal “*que-darían baleiros de contido real outros dereitos que a Constitución consagra, reducidas a formas ocas as institucións representativas e absolutamente falseado o principio de liberdade democrática que enuncia o art. 1 ap. 2º CE e que é a base da nosa organización xurídico-política*”²².

Como garantía desta opinión pública libre, o Tribunal sinala as liberdades do artigo 20 da Constitución, de tal modo que garantir as liberdades de expresión e información, permitíndolle á poboación acceder libremente a información plural e veraz, supón tamén garantir a libre formación da opinión pública, pois baseándose nesta información poderán formar de xeito responsable a súa propia opinión sobre os distintos asuntos en atención aos feitos²³.

Hoxe en día, 40 anos despois de que o TC escribise esas palabras, a poboación ten máis acceso que nunca á información; de feito, nunca tivemos acceso a tal cantidade de información con tan pouco esforzo e, porén, non parece que se poida afirmar que estamos mellor informados, e iso débese en gran parte á calidade da información que temos á nosa disposición, por distintos motivos.

O primeiro é que, fronte ao sistema anterior, en que a información proviña con carácter principal dos medios de comunicación, suxeitos a responsabilidade no caso de que difundisen informacións falsas sen ter despregado un dilixente labor de contrastación, a día de hoxe calquera pode ser comunicador²⁴, calquera pode subir informacións á rede nunha sorte de “xornalismo cidadán”²⁵, sen máis límites que os

²⁰ Define Antonio Torres del Moral a opinión pública como o precipitado de ideas, crenzas e sentimentos que se expresa na vida pública, xa que, sen expresión externa, a opinión perde o seu carácter de pública e deixa de cumprir as importantes funcións que a caracterizan. TORRES DEL MORAL, A. (dir.), *Libertades informativas*, Colex, 2009, p. 140.

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

²² Sentenza do Tribunal Constitucional 12/1982, do 31 de marzo, FX 3.º.

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

²⁴ Así o sinala o Tribunal Constitucional na Sentenza 8/2022, do 27 de xaneiro, FX 2.º, cando di que “*o uso das ferramentas dixitais converte os seus usuarios en creadores de contidos, emisores, difusores e reprodutores deses contidos (...). Polo tanto, os usuarios poden chegar a desempeñar un papel moi próximo ao que viñan desenvolvendo ata agora os xornalistas nos medios de comunicación tradicionais*”.

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

establecidos para a liberdade de expresión, que necesariamente deben ser mínimos nun sistema democrático²⁶.

Esta proliferación de comunicadores non tería sido posible sen o desenvolvemento das novas tecnoloxías e o ascenso das redes sociais e de mensaxaría como principais formas de acceso á información e ás noticias²⁷. Hoxe en día, gran parte da poboación non se informa a través dos medios tradicionais, nin sequera da prensa dixital, senón que o fan a través do seu *feed* de Facebook, X, ou a rede social que cada un siga²⁸, e que os medios de comunicación utilizan como altofalante para as súas publicacións²⁹, a miúdo facendo uso de titulares sensacionalistas ou *clickbait*³⁰, que podemos considerar desinformación en si mesmos, xa que non achegan nada ao xornalismo, só á conta de resultados.

É a rede e o seu algoritmo a que decide que contido lle mostra a cada persoa usuaria, controlando polo tanto o fluxo da información, sen que a cidadanía poida coñecer os nesgos ideolóxicos ou intereses detrás da rede debido á falta de transparencia dos seus algoritmos³¹.

O ascenso das redes sociais como propagadores de noticias viuse impulsado por distintos motivos, entre os que destacan a inmediatez na difusión, o acceso simultáneo a unha grande audiencia, a accesibilidade das redes a calquera persoa (tanto para publicar como para recibir contidos) e, aínda que moitas veces as propias persoas usuarias non sexan conscientes, a facilidade para establecer filtros e segmentar con gran detalle usuarios para fins concretos³².

Esta segmentación fai que recibamos información “cociñada” algorítmicamente en vez de en atención á relevancia da noticia ou mesmo a unha determinada liña editorial, o que pode afectar á calidade da propia información³³ e, o que é aínda máis importante, ao proceso de formación da opinión pública, xa que esta segmentación ou microsegmentación pode ser maliciosamente utilizada con fins desinformativos.

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*. EDERSA Editoriales de Derecho Reunidas, 1991, p. 69.

²⁷ Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C) (2023). *Op. cit.*, p. 7/26.

²⁸ Sobre isto afonda Emily Bell no seu artigo “Facebook se está comiendo el mundo”. *Cuadernos de Periodistas*, 32, 2016, pp. 67–76. Dispoñible en <https://www.cuadernosdeperiodistas.com/facebook-se-esta-comiendo-mundo/>

²⁹ JIMÉNEZ PALOMARES, P. e 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 Seguindo a Berta García e outros, podemos definir o *clickbait* como “o titular dunha noticia que non responde aos criterios xornalísticos tradicionais e cuxo obxectivo final é manter o receptor na páxina o maior tempo posible, non informar. En concreto, o titular cebo tería como obxectivo principal a comercialización ou difusión da información, mentres que o titular xornalístico mantería o rol principal de informar os usuarios”. GARCÍA OROSA, B., GALLUR SANTORUN, S. e 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 a microsegmentación, unha práctica publicitaria que permite dirixir contidos a grupos específicos baseándose nos seus intereses, características, etc., personalizando as mensaxes que reciben. Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C) (2023). *Op. cit.*, p. 14/26.

³³ Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C) (2023). *Op. cit.*, p. 8/26.

En segundo lugar, porque neste ámbito en que calquera pode publicar contidos e replicalos sen contrastar o máis mínimo a súa veracidade, e en que podemos estar conectados a centos de persoas a través das redes sociais, intensifícanse os efectos dos nesgos cognitivos³⁴, especialmente do nesgo de confirmación, definido como a tendencia dunha persoa a favorecer e priorizar a información que confirma as súas crenzas ou hipóteses.

É dicir, de entre a cantidade inxente de información e informadores aos que podemos acceder, tenderemos a seleccionar ou priorizar as opinións dos que confirmen as nosas, creando unha sorte de “burbullas de filtro”³⁵ ou “cámaras de eco ou resonancia”³⁶ onde o propio grupo cada vez está máis convencido de estar no certo, pois practicamente toda a información que recibe reforza as súas crenzas e rexeita as contrarias³⁷.

Acontece ademais que para o noso cerebro resulta moi difícil convivir con incongruencias, polo que tratamos de encher canto antes os ocos que quedan nos nosos razoamentos, mesmo aínda que sexa con información falsa³⁸. Preferimos autoenganarnos e gozar da harmonía da coherencia entre os nosos actos e crenzas (ou entre as nosas crenzas entre si) a ter que enfrontarnos á ansiedade que supón convivir con dúas crenzas en conflito ou co conflito entre os nosos pensamentos e o noso comportamento, o que Leon Festinger bautizou como *disonancia cognitiva*³⁹.

En terceiro lugar, e aquí reside probablemente a maior ameaza, pola irrupción das ferramentas de intelixencia artificial que permiten, por un lado, un uso moito máis preciso, eficiente e masivo dos datos dos electores, como aconteceu nas eleccións presidenciais estadounidenses de 2016 con Cambridge Analytica. Por outro lado, permítenos construír unha realidade á medida, fabricando os datos, imaxes, vídeos

³⁴ Sobre os nesgos cognitivos, GARRIGUES WALKER, A. e GONZÁLEZ DE LA GARZA, L. M., *El Derecho a no ser engañado. Y cómo nos engañan y nos autoengañamos*, Aranzadi, 2020, pp. 85 e ss., seguindo a Stuart Sutherland, definen o nesgo cognitivo como “un fenómeno psicológico de natureza involuntaria, xa que se produce de forma automática, e que describe unha alteración na mente humana que a conduce a unha distorsión da percepción, a unha distorsión cognitiva e ou a xuízos imprecisos ou a interpretacións ilóxicas ou irracionais dos feitos”.

³⁵ As burbullas de filtro son un fenómeno polo que unha persoa se atopa exposta principalmente a información e perspectivas que reforzan as súas crenzas e puntos de vista preexistentes, debido a algoritmos de recomendación e personalización en plataformas dixitais e redes sociais, todo iso sen que se requira un papel activo por parte do usuario. Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C) (2023). *Op. cit.*, p. 13/26.

³⁶ As cámaras de resonancia son un ámbito ou plataforma onde as ideas, mensaxes ou conceptos atopan unha audiencia receptiva e amplifican o seu impacto, requirindo dun papel activo do usuario, que demanda activamente determinados contidos por diferentes razóns, incluídas as ideolóxicas, e onde pode circular máis libremente a desinformación. Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (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. e YEZERS'KA, L., 2022, *op. cit.*, p. 155; Oficina de Ciencia e Tecnoloxía do Congreso dos Deputados (Oficina C) (2023). *Op. cit.*, p. 13/26; BERROCAL-GONZALO, S., WAISBORD, S. e 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. e GONZÁLEZ DE LA GARZA, L. M., 2020, *op. cit.*, p. 64.

³⁹ Leon Festinger considera que a disonancia, a existencia de relacións entre coñecementos, opinións, crenzas ou condutas que non concordan, é un factor de motivación, xa que condiciona a nosa conduta no sentido de dirixila a reducir esa disonancia. Extractado por A OVEJERO a partir da obra de L. Festinger: *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.

ou audios necesarios para sustentar esa creación e facéndoo de tal forma que resulten indistinguibles da propia realidade, o que se coñece como *deepfakes*⁴⁰.

Un dos problemas clave das *deepfakes* é que atacan os propios principios da estrutura social en que a maioría da sociedade asume que hai algunha forma de verdade mutuamente aceptada, como eran ata agora os contidos audiovisuais, polo que a irrupción dunha tecnoloxía que rompe con ese acordo social ou, polo menos, o pon en dúbida, supón unha perda de confianza epistémica⁴¹ que deriva na fin da crenza na verdade⁴². E isto supón non só que poidamos tomar por certo un contido fabricado, senón tamén que poidamos ter por falso un audio, vídeo ou imaxe real⁴³.

Polo tanto, atopámonos intentando sobrevivir nun océano de información, na súa maioría non contrastada, onde convive información veraz con información terxiversada ou sacada de contexto, parodias, información falsa, información atribuída a fontes incorrectas, información fabricada... e é baseándose naquilo que seleccionemos deste inmenso océano como formaremos a nosa opinión sobre os distintos temas e, polo tanto, como se configurará a opinión pública.

Neste escenario, para calquera cos medios axeitados resultaría relativamente sinxelo influír na formación da opinión pública, orientando os electores nunha dirección ou noutra alterando o sentido do seu voto e influíndo, polo tanto, de xeito claro no resultado dos procesos electorais, coas implicacións que iso pode ter mesmo a nivel de seguridade nacional, especialmente cando quen manexa os fíos é unha potencia estranxeira⁴⁴.

Para intentar controlar este risco, a Unión Europea creou dúas comisións sucesivas de Inxerencias Estranxeiras, a primeira en 2020 e a segunda en 2022. Froito do traballo destas comisións, o 1 de xuño de 2023 o Parlamento Europeo aprobou a Resolución P9_TA(2023)0219 sobre as inxerencias estranxeiras en todos os procesos democráticos da Unión Europea, en particular a desinformación.

Nesta resolución, que busca acordar unha estratexia común para a Unión na loita contra as inxerencias estranxeiras, sinálase que estas poden adoptar diversas formas: desde fraude nas papeletas ou bloqueo dos colexios electorais ata a distribución de información distorsionada sobre os candidatos ou o propio proceso electoral e

⁴⁰ Se ben os primeiros usos da tecnoloxía *deepfake* se atopan no ámbito da pornografía, a primeira documentación desta tecnoloxía encontrámola no ano 2017, cando tres investigadores da Universidade de Washington demostraron como se podía manipular a tecnoloxía de edición de imaxes dinámicas para crear novos contidos, concretamente imitando a voz e expresións faciais do presidente Obama. SUPASORN SUWAJANAKORN, STEVEN M. SEITZ, AND IRA KEMELMACHER-SHLI-ZERMAN. *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., e 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, p. 6.

⁴⁴ Segundo os datos do Eurobarómetro Flash n.º 522 (2023) sobre "Democracia", o 43% dos europeos mencionou como unhas das ameazas máis graves para a democracia a propaganda ou a información falsa ou enganosa procedente de fontes estranxeiras non democráticas e a inxerencia estranxeira encuberta na política e na economía do seu país.

campañas de desinformación⁴⁵, pasando por financiamento a determinadas opcións políticas ou apoio a correntes ideolóxicas extremas.

En opinión do Parlamento, o obxectivo destas inxerencias é distorcer a política nacional para socavar ou destruír a confianza a longo prazo da cidadanía na lexitimidade das súas institucións e procesos democráticos⁴⁶, constituíndo non só unha ameaza á libre formación da opinión pública, senón tamén á seguridade nacional e transfronteiriza⁴⁷.

3 A manipulación do electorado a través de propaganda cognitiva e *microtargeting*: o caso de Cambridge Analytica

A propaganda cognitiva ou mercadotecnia cognitiva é un tipo de publicidade que se centra en aproveitar os sentimentos ou a emoción para conseguir unha reacción na persoa destinataria. Estúdase o usuario a fondo, os seus gustos, os contidos que publica, o que comparte, buscando aliñar a marca en cuestión con todo aquilo que lle gusta ao consumidor ou que este considera positivo ou desexable, de tal forma que se sinta “naturalmente” atraído cara a ela. O novidoso é que non é unha propaganda pasiva, senón que se aproveita dos nesgos caracterolóxicos das persoas destinatarias para deseñar unha campaña de gran precisión á medida de cada persoa e das súas preferencias emocionais e políticas⁴⁸.

Antes da expansión das redes sociais, este tipo de publicidade resultaba case impensable, xa que por grande que fose o esforzo de coñecer o usuario a través de enquisas ou do estudo das súas compras, nunca houbo tanta facilidade como agora para acceder a toda esa información e procesala.

O mercado tecnolóxico ofrécenos multitude de aplicacións e servizos supostamente gratuítos que instalamos nos nosos dispositivos e comezamos a usar facilitando todos os datos que o servizo en cuestión solicita para configurar o noso perfil e así mellorar a experiencia de usuario mediante a personalización dos contidos. Todo iso

⁴⁵ Resolución P9_TA(2023)0219 do Parlamento Europeo, do 1 de xuño de 2023, sobre as inxerencias estranxeiras en todos os procesos democráticos da Unión Europea, en particular a desinformación (2022/2075 (INI)), Considerando AZ.

⁴⁶ Resolución P9_TA(2023)0219 do Parlamento Europeo, do 1 de xuño de 2023, sobre as inxerencias estranxeiras en todos os procesos democráticos da Unión Europea, en particular a desinformación (2022/2075 (INI)), Considerando BA.

⁴⁷ A Axencia da Unión Europea para a ciberseguridade (ENISA) e o Servizo Europeo de Acción Exterior (EEAS) falan de inxerencia e manipulación estranxeira da información (FIMI, pola súa sigla en inglés) para referirse a un patrón de comportamentos, principalmente non ilegais que ameaza ou ten o potencial de impactar negativamente nos valores, procedementos e procesos políticos. Trátase dunha actividade manipulativa por natureza, levada a cabo de xeito intencionado e coordinado, cuxos actores poden ser tanto actores estatais como non estatais, incluídos os seus representantes dentro e fóra do seu propio territorio. ENISA, EEAS. *Foreign Information Manipulation Interference (FIMI) and Cybersecurity - Threat Landscape*. DOI: 10.2824/750. (2022) Dispoñible 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.

sen sermos conscientes ou menosprezando o feito de que, como acontece noutros ámbitos, se non nos cobran polo produto, é que o produto somos nós⁴⁹.

Na rede queda rexistro do noso perfil, idade, sexo, situación, intereses declarados... pero tamén de cada unha das nosas accións e interaccións dentro da propia rede: que é o que publicamos, que contas ou páxinas seguimos, que contidos nos gustan ou replicamos, mesmo cal é a nosa reacción ante determinadas publicacións, anuncios, noticias... Queda constancia de cales son os nosos medios de comunicación ou tendencias políticas afíns a través das noticias que publicamos, comentamos ou replicamos, pero tamén a través dos políticos que seguimos ou criticamos... en resumo, toda a nosa actividade a través das redes sociais queda rexistrada e perfectamente segmentada por sexo, ideoloxía, situación, franxa de idade, etc., para ser usada da forma que máis lle conveña ao mellor ofertante.

Precisamente isto é o que aconteceu en 2016 durante as eleccións presidenciais estadounidenses, nas cales a empresa Cambridge Analytica, unha rama do Grupo SCL⁵⁰, contratada polo candidato Donald Trump, utilizou indebidamente datos persoais de millóns de usuarios de Facebook para crear perfís psicolóxicos concretos dos usuarios e así deseñar anuncios políticos co ton e contido axeitados para cada perfil co obxectivo de dirixir o seu sentido do voto, á vez que se creaban e difundían noticias falsas que aumentaban o efecto buscado. Donald Trump gañou as eleccións e non podemos saber cal sería o resultado de non ter recorrido a este tipo de prácticas.

Non estamos a falar da propaganda electoral habitual, senón de publicidade feita á medida de cada usuario utilizando perfís psicométricos, que busca remover as súas emocións para conseguir unha actuación pola súa parte⁵¹. O obxectivo xa non é convencer o usuario para que vote a un candidato determinado, senón xerar nel a emoción que provoca que se convenza só. Polo tanto, a propaganda cognitiva non apela á razón, senón ás emocións, á alegría, á rabia, á ilusión e á vinganza... de modo que o propio destinatario non chega a ser consciente de que está a ser influenciado, de que verdadeiramente non está a decidir libremente.

Para conseguir os datos, recorreron a un académico da Universidade de Cambridge, Aleksandr Kogan, que ideou unha aplicación en liña chamada *This is your digital life* que se presentaba como un test en liña e afirmaba ser unha ferramenta de investigación, que mesmo remuneraba simbolicamente as persoas participantes. Para finalizar o test, os usuarios debían iniciar sesión en Facebook e outorgarlle certos permisos á aplicación, como acontece en tantos outros casos⁵².

⁴⁹ Sobre este tema: MARTÍN-HERRERA, I. e 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) é unha empresa británica fundada en 1990, dedicada á investigación do comportamento e comunicación estratéxica.

⁵¹ Resulta moi didáctica a explicación de BYUNG-CHUL HAN sobre os efectos do *microtargeting* no seu libro *Infocracia*: "No *microtargeting*, os votantes non están informados do programa político dun partido, senón que se lles manipula con publicidade electoral adaptada ao seu psicograma, e non poucas veces con *fake news*. (...) Os cidadáns deixan de estar sensibilizados para as cuestións importantes, de relevancia social. (...) Deste xeito, socavan un principio fundamental da democracia: a *autoobservación da sociedade*." HAN, B., 2022, *op. cit.*, p. 37.

⁵² ISAAK, J. e 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.

Aproximadamente 270 mil persoas cubriron a enquisa e concederon os permisos que a aplicación solicitaba, crendo que os seus datos só serían utilizados no seo da investigación e con fins académicos. Pero entre os permisos que a aplicación solicitaba estaba o acceso aos datos dos amigos na rede social, o que lle permitiu á empresa acceder aos datos duns 87 millóns de persoas⁵³.

Conseguidos os datos, Cambridge Analytica serviuse de ferramentas de *big data*⁵⁴ e *machine learning*⁵⁵ para tratalos en masa, cruzando as respostas das enquisas cos likes e publicacións compartidas en Facebook polos distintos usuarios. Isto permitiulles segmentar os 87 millóns de usuarios en pequenos grupos de persoas ou perfís psicográficos⁵⁶ preocupadas por asuntos concretos e que, polo tanto, serían receptivas a publicacións sobre esas materias concretas se se utilizaba o ton axeitado.

Esta minaría de datos, que con medios tradicionais tería sido imposible, convértese en algo sinxelo e rápido grazas á intelixencia artificial e, concretamente, á *machine learning*, que permite obter en tempo récord unha segmentación extremadamente escurpulososa e detallada.

Unha vez segmentados os votantes nas distintas categorías, utilizouse un sistema de intelixencia artificial capaz de analizar centos ou mesmo miles de variacións dun anuncio para comprobar cal era a que mellor podería funcionar con cada un dos perfís psicográficos previamente identificados. Tratábase de personalizar ata o máis mínimo detalle de cada anuncio, desde o tipo e cor de letra ata o ton da mensaxe, pasando pola iconografía ou a música, no caso do contido audiovisual.

Polo tanto, a cada pequeno grupo mostrábaselle só a variación do anuncio que o sistema detectara que mellor funcionaría, chegando a deseñarse entre 40 e 50 mil variantes de diferentes argumentos electorais⁵⁷, que se adaptaban á perfección ao perfil do usuario ao que se lle amosaban, sen que ademais puidesen ser coñecidos por outros usuarios, xa que se basean en publicacións invisibles ou *dark post*⁵⁸, o que fai que tamén sexan dificilmente rastrexables polos organismos de control.

É dicir, un usuario non pode ver este tipo de publicacións no muro doutro usuario, senón que só se mostran ao seu destinatario, aumentando así a personalización e

⁵³ Nun primeiro momento pensouse que Cambridge Analytica accedera aos datos de aproximadamente 50 millóns de usuarios, pero os datos en 1 de maio de 2018 proporcionados por Meta, propietaria de Facebook, apuntan a 87 millóns de persoas, a maioría delas en Estados Unidos (81,6%). Datos dispoñibles na web: <https://about.fb.com/news/2018/04/restricting-data-access/>

⁵⁴ O analista Dan Kusnetzky define *big data* como as ferramentas, procesos e procedementos que lle permiten a unha organización crear, manipular e administrar grandes conxuntos de datos e instalacións de almacenamento. CAMARGO ORTEGA, J. F., CAMARGO VEGA, J. J., e JOYANES AGUILAR, L., "Conociendo Big Data", *Revista FI-UPTC*, 24(38), 2015, pp. 63-77, p. 63.

⁵⁵ *Machine learning* ou aprendizaxe automática é unha rama da intelixencia artificial que lles permite ás computadoras aprender dos datos sen seren programadas explicitamente para cada tarefa. En lugar de seguir instrucións fixas, os sistemas de *machine learning* analizan patróns nos datos, o que lles permite facer predicións ou tomar decisións baseadas na información que procesaron.

⁵⁶ Defínese a psicografía como o estudo e a clasificación das persoas segundo as súas actitudes, aspiracións e outros criterios psicolóxicos.

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

⁵⁸ Son publicacións cuxa visibilidade se condiciona á pertenza a determinados grupos seleccionados polo administrador da páxina. Trátase dunha funcionalidade especialmente pensada para o mundo publicitario que permite non saturar o mercado mostrando información que poida non resultar relevante para grupos determinados, por exemplo por idade ou situación, e dirixirse soamente á audiencia á que a publicación pode interesar.

evitando que o usuario se decate de que está sendo manipulado, ao tempo que se evita ou atrasa un debate público e aberto sobre o contido e a veracidade desas publicacións que permitiría contrastalas⁵⁹.

Ademais de todo o exposto, debemos ter en conta que a propaganda cognitiva non opera de xeito illado. Non se trata só de que se lle mostren ao usuario anuncios especialmente deseñados para el en función dos seus intereses, previamente conseguidos a través das cesións de datos efectuadas, consciente diso ou non, polo propio usuario, senón que a miúdo se combina esta ferramenta con outras como o uso de contas falsas automatizadas⁶⁰ en redes sociais, tamén coñecidas como *social bots*⁶¹.

No caso das eleccións presidenciais estadounidenses de 2016, o uso de contas automatizadas foi estratéxico e deliberado, especialmente no lado republicano, axustando á perfección o momento de produción de contido durante os debates e chegando no día das eleccións a superar os *bots* pro-Trump os de Clinton a razón de 5 a 1⁶².

Estas contas automatizadas utilízanse sobre todo para alterar o debate na propia rede, dando audiencia a publicacións que poderían parecer inicialmente irrelevantes, chegando a viralizar contidos que doutro xeito pasarían desapercibidos, apropiándose dun debate creado inicialmente por humanos ou mesmo creando *ex novo* un debate que doutra forma non existiría⁶³. Todo iso cun volume de publicacións e interaccións moi superior ao humanamente posible, o que consegue con pouco esforzo manipular o peso do apoio ás distintas opcións, inclinando de maneira artificial a balanza cara a unha delas que aparece claramente como gañadora⁶⁴.

Todo isto consegue reforzar a burbulla do usuario e con ela a polarización existente, unha polarización en que os extremos xa non son tan mal vistos⁶⁵, pois hai gran cantidade de contas que defenden as súas posicións á calor do anonimato das redes e que, ademais, atopan explicación ao inexplicable, normalmente baseándose nalgunha retorcida conspiración que casualmente o explica todo⁶⁶.

Todo iso amplificado polo efecto da disonancia cognitiva: necesitamos buscar explicacións que reforcen as nosas crenzas e isto fai case imposible, a medida que se

⁵⁹ ANDERSON, B. e 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 Unha conta automatizada ou *bot* é un software ou programa informático elaborado para levar a cabo determinados tipos de actividades repetitivas como se dun ser humano se tratase; pode xerar contido, replicalo, seguir ou deixar de seguir contas, enviar mensaxes directas... todo iso imitando o comportamento 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* é unha conta controlada por software que participa activamente nas redes sociais para influir na opinión pública na dirección desexada: 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. e 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.

⁶⁴ Calcúlase que na campaña española das eleccións xerais de novembro de 2019 había máis de 40.000 *bots* tuiteando sobre eventos destacados da 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 a polarización: BERROCAL-GONZALO, S., WAISBORD, S. e GÓMEZ-GARCÍA, S., 2023, *op. cit.*

⁶⁶ SÁDABA RODRÍGUEZ, I. e 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.

avanza neste tobo, dar marcha atrás, rectificar os nosos pensamentos, pois preferiremos crer calquera cousa antes que rexeitar todas as teorías que xa defendemos e que, polo menos no contexto da nosa burbulla, encaixan á perfección. A credibilidade da información é directamente proporcional a canto coincida cos nosos valores, crenzas e prexuízos⁶⁷.

Resulta importante, non obstante, aclarar que o problema non é o uso das redes sociais para chegar ao electorado –de feito, a campaña doutro candidato á presidencia de Estados Unidos, Barack Obama, foi das primeiras en explotar as redes sociais, en concreto Facebook e MySpace⁶⁸ para facer chegar a súa mensaxe–, a clave é que o que facía era explotar as ferramentas que a rede social puña ao seu servizo para chegar ao electorado a través desa vía⁶⁹.

Pola contra, no caso da campaña de Donald Trump, o que realmente fixo Cambridge Analytica foi enganar os usuarios para que lles confiasen os seus datos, a súa información, e utilízala con fins totalmente distintos aos declarados. Isto supón unha clara violación da debida protección de datos das persoas usuarias, que resulta aínda de maior gravidade porque se produce con fins electorais.

O obxectivo é modificar o sentido do voto dos usuarios da rede escudriñando na súa privacidade para utilizar os datos roubados co obxectivo de lle facer crer a ese usuario que é el quen está a elixir, cando realmente o que ocorre é que se condicionou a súa vontade a un nivel profundo, mostrando só os contidos que reafirmaban a postura que interesaba e creando propaganda á medida do usuario aproveitando a información roubada.

A propaganda computacional sitúase, polo tanto, como unha poderosa ferramenta contra a democracia⁷⁰, ao afectar ao sentido de voto a través da manipulación da formación da opinión pública⁷¹.

Evidentemente, pode que o resultado das eleccións estadounidenses de 2016 tivese sido o mesmo sen a actuación de Cambridge Analytica, pero é imposible sabelo. O que si sabemos é que un dos candidatos presidenciais levou a cabo unha campaña pouco transparente, roubando e tratando os datos de millóns de persoas e utilizando propaganda cognitiva para asegurar o seu voto e que esta situación acabou supoñendo a quebra de Cambridge Analytica e grandes perdas para a rede social Facebook, mentres que o candidato que contratou os servizos da consultora e gañou as eleccións seguiu no seu posto ata rematar o seu mandato.

⁶⁸ BARRIENTOS-BÁEZ, A., CALDEVILLA-DOMÍNGUEZ, D. e YEZERS'KA, L., 2022, *op. cit.*, p. 158.

⁶⁹ Ao celebrarse as eleccións, Obama reunía 3 millóns de seguidores en Facebook, 700.000 en MySpace, 100.000 en Twitter e 144.615 subscritores 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. e 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/index.29.1.6588>, p. 109.

⁷⁰ WOOLLEY, S. C. e 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.

4 O uso da intelixencia artificial para atacar os candidatos: o caso de Eslovaquia

As ferramentas que ofrece a intelixencia artificial, concretamente no eido da desinformación, supoñen un reto nunca visto. No apartado anterior afondouse nalgún dos usos das ferramentas de *machine learning* neste campo, pero non podemos rematar sen tratar o papel das ferramentas *deep learning*⁷², que permiten crear contidos audiovisuais indistinguibles da realidade, ata o punto de poder utilizar a imaxe ou a voz dunha persoa real para que faga ou diga cousas que xamais fixo ou dixo.

Podemos imaxinar infinidade de formas en que esta tecnoloxía pode afectar ao proceso de toma de decisións en materia electoral, pero o certo é que xa temos exemplos reais.

O 30 de setembro de 2023 celebráronse eleccións parlamentarias anticipadas en Eslovaquia nas cales o ex primeiro ministro prorruso Robert Fico, líder de Dirección-Socialdemocracia Eslovaca (SMER-SSD)⁷³, e Michal Šimečka, líder do partido liberal Eslovaquia Progresista (EP), disputaban o posto de primeiro ministro.

Apenas dous días antes da celebración dos comicios, publicouse en Facebook un audio duns dous minutos en que supostamente se escoitaba o candidato de EP discutindo coa xornalista Monika Tódová a forma de amañar o proceso electoral comprando votos da minoría xitana. Para non deixar lugar a dúbidas, o audio estaba subtítulado e ía acompañado da imaxe estática do candidato liberal e da xornalista. Ambos saíron rapidamente a desmentir o audio, pero con certa moderación, dado que o país se atopaba no período de reflexión, que no caso eslovaco se estende durante as 48 horas previas ao comezo da votación⁷⁴.

Tamén o departamento de comprobación de información da axencia de noticias AFP⁷⁵ saíu rapidamente a pór de manifesto que a gravación mostraba signos de ter sido manipulada con intelixencia artificial⁷⁶.

⁷² Define Raúl López Briega a *deep learning* ou aprendizaxe profunda como un subcampo dentro da *machine learning* que utiliza distintas estruturas de redes neuronais para lograr a aprendizaxe de sucesivas capas de representacións cada vez máis significativas dos datos. A aprendizaxe profunda ou *deep learning* fai referencia á cantidade de capas de representacións que se utilizan no modelo; en xeral adóitanse utilizar decenas ou mesmo centos de capas de representación que aprenden automaticamente a medida que o modelo é adestrado cos datos. Citado en GUAÑA-MOYA, J., SALGADO-REYES, N., e 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 Información*, 2022, pp. 116-127. Dispoñible 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>.

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

⁷⁵ A Agence France-Presse (AFP) é unha das tres principais axencias mundiais de información, que conta ademais cun departamento de verificación de noticias falsas.

⁷⁶ Údajná nahrávka telefonátu predsedu PS a novinárky Denníka N vykazuje podľa expertov početné známky manipulácie (A suposta gravación da chamada entre o presidente do PS e a xornalista de Denník N mostra, segundo os expertos, múltiples signos de manipulación). AFP FACTY. 28 de setembro de 2023 [consulta 15 de outubro de 2024] Dispoñible en <https://fakty.afp.com/doc.afp.com.33WY9LF>.

No entanto, sen poderen ofrecer os supostos participantes probas contundentes sobre a súa falsidade (sen dúbida unha *probatio diabólica*⁷⁷), o audio seguiu en liña. Na política de recursos manipulados de Meta contemplábase a posibilidade de eliminar vídeos falsos onde aparecen persoas dicindo cousas que nunca dixeron, pero as ditas directrices non incluían expresamente os audios, polo que a gravación se atopaba nunha sorte de lagoa respecto á política da compañía.

E non foi este o único uso da intelixencia artificial durante este proceso electoral. Tres días antes das eleccións, publicárase outro audio, tamén do candidato liberal Michal Šimečka, en que afirmaba que duplicaría o prezo da cervexa se gañaba as eleccións co obxectivo de loitar contra o alcoholismo. Nese caso a propia empresa Meta alertou da existencia do audio, o que puxo os verificadores a traballar, conseguindo probar nunhas horas que o contido efectivamente era falso, tal como declarara o seu protagonista⁷⁸.

Á marxe da influencia que puidesen ter os audios sobre os resultados do proceso electoral, que perdeu o candidato de Eslovaquia Progresista en contra do que as enquisas prevían⁷⁹, considero importante destacar tres feitos.

O primeiro deles refírese ao problema para exixir responsabilidades respecto dos danos creados co uso de tecnoloxías *deep learning*. A tecnoloxía utilizada fai moi difícil en primeiro lugar detectar que se trata de contido falso ou manipulado e, en segundo lugar, detectar a verdadeira orixe dos contidos, o que fai imposible a rendición de contas; de feito, hoxe en día non se coñece a orixe do audio *deep fake* sobre o candidato de EP, co problema que iso supón en canto á exixencia de responsabilidades.

Isto cobra especial relevancia, e conecto xa co segundo dos feitos que cómpre destacar, no caso de que quen queira manipular o proceso electoral sexa unha potencia estranxeira. Xa no verán de 2023 a Unión Europea puxo de manifesto a súa preocupación pola posibilidade de inxerencias estranxeiras nas eleccións no territorio da Unión afirmando que “a desinformación reduce a capacidade dos cidadáns para tomaren decisións informadas e participar libremente nos procesos democráticos” e que “algúns Estados membros, que Rusia considera a súa esfera de influencia, están máis expostos a riscos xeopolíticos derivados da inxerencia do Kremlin no seu espazo de información”⁸⁰.

De xeito case premonitorio, esta resolución sinalaba como posible inxerencia nos procesos electorais a distribución de información distorsionada sobre os candidatos e as campañas de desinformación en redes sociais. É dicir, sabíase que podía suceder e, porén, non puido evitarse nin a difusión das loiadas e *deepfakes* nin as súas consecuencias.

⁷⁷ Tamén chamada proba inquisitorial, a proba diabólica é aquela pola que se lle exige á persoa acusada que probe que algo non existe, é dicir, que probe un feito 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 outubro de 2023 [consultado o 15 de outubro de 2024]. Dispoñible en <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/>

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

⁸⁰ Resolución P9_TA(2023)0219 do Parlamento Europeo, do 1 de xuño de 2023, sobre as inxerencias estranxeiras en todos os procesos democráticos da Unión Europea, en particular a desinformación (2022/2075 (INI)), Considerando T.

O terceiro dos aspectos que debemos resaltar é o feito de que a interlocutora do audio *deepfake* fose unha xornalista, o cal resulta importante por varios motivos. O primeiro deles é que nos atopamos ante unha persoa que non é unha personaxe pública e, a pesar diso, vese envolta de xeito involuntario e forzado nun asunto de corrupción electoral, co dano que iso pode supor para a súa honra e a súa imaxe a nivel tanto persoal como profesional.

O segundo motivo é que, ademais, esta interlocutora é unha persoa que se dedica profesionalmente ao xornalismo, resultando moi difícil pensar que esta elección fose azarosa, senón que máis ben parece unha opción premeditada que busca non só desacreditar o candidato, quen pola súa dedicación á cousa pública debe soportar un volume maior de inxerencias sobre a súa honra e propia imaxe, e sementar as dúbidas sobre a lexitimidade do proceso electoral, senón tamén pór en dúbida a xornalista en particular e o xornalismo no seu conxunto⁸¹.

Neste punto, debemos lembrar o lugar clave que lle outorga o TEDH á prensa nos sistemas democráticos, protexendo o seu papel na formación da opinión pública mediante a publicación de información sobre asuntos de interese público veraz e contrastada, obrigando as autoridades públicas e outras esferas de poder a render contas e sinalando a corrupción e o abuso de poder⁸². Ao involucrar unha xornalista nunha conspiración para cometer fraude electoral, estase a atacar tamén o papel da prensa como vixilante público⁸³, sementando a dúbida sobre a súa imparcialidade e o seu imprescindible papel na formación da opinión pública.

5 Conclusións

As ferramentas de intelixencia artificial supoñen unha revolución tecnolóxica sen precedentes. Estamos ante o maior progreso tecnolóxico do noso tempo, equiparable ou mesmo de maior relevancia que o que supuxo no seu día a máquina de vapor ou a electricidade⁸⁴, unha revolución con efectos transversais en todas as áreas do coñecemento e da ciencia e que nos sitúa nun novo escenario tecnolóxico, produtivo e social⁸⁵.

Non obstante, resulta extremadamente importante avaliar e prever os riscos aos que pode someternos a IA, pois estamos ante tecnoloxías disruptivas que, usadas maliciosamente, teñen a potencialidade de causar danos de consecuencias catastróficas, pondo en xaque a democracia e ata a propia humanidade.

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

⁸³ Entre outros, asunto Pedersen and Baadsgaard c. Denmark STEDH do 17 de decembro 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.

Ao longo deste traballo exploráronse algúns dos usos da intelixencia artificial que xa se manifestaron como ameazas reais para a libre formación da opinión pública e, polo tanto, para o sistema electoral e democrático, que se asenta no presuposto de que a poboación acode a votar tras ter formado a súa opinión de xeito libre e baseándose en información veraz e de calidade⁸⁶.

No caso da microsegmentación, atopámonos en primeiro lugar cun problema en materia de protección de datos persoais, dado que a miúdo este tipo de prácticas peca de falta de permisos na recollida ou tratamento dos datos. Interesa aclarar que no seo da Unión Europea é un aspecto que xa se está a regular, sen que polo momento poidamos sacar conclusións sobre a efectividade desa regulación.

Por un lado, o Regulamento xeral de protección de datos⁸⁷, malia non regular especificamente a microsegmentación, si a limita a través das restricións xerais na recollida e tratamento dos datos e no establecemento de certos principios no tratamento dos datos⁸⁸.

Si regula de xeito específico a segmentación en materia electoral o recente Regulamento sobre transparencia e segmentación en publicidade política⁸⁹, cuxo considerando 6 advirte de que a microsegmentación e outras técnicas avanzadas de tratamento dos datos “poden supor graves ameazas para intereses públicos lexítimos como a equidade, a igualdade de oportunidades e a transparencia nos procesos electorais e nos dereitos fundamentais de liberdade de expresión, privacidade e protección dos datos persoais, así como de igualdade e non discriminación e no dereito fundamental a ser informado de xeito obxectivo, transparente e plural”.

Este regulamento permite a segmentación sempre que se cumpran tres requisitos: 1) que o interesado lle proporcionase os datos persoais ao responsable do tratamento; 2) que o interesado dese, por separado, o seu consentimento expreso ao tratamento dos datos persoais para os efectos de publicidade política, e 3) que as ditas técnicas non impliquen a «elaboración de perfís»⁹⁰.

Por último, o Regulamento de intelixencia artificial⁹¹ prohibe expresamente no seu artigo 5.1.a) baixo pena de sanción de ata 35 millóns de euros (a concretar polos

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

⁸⁷ REGULAMENTO (UE) 2016/679 DO PARLAMENTO EUROPEO E DO CONSELLO do 27 de abril de 2016 relativo á protección das persoas físicas no que respecta ao tratamento de datos persoais e á libre circulación destes datos e polo que se derroga a Directiva 95/46/CE (Regulamento xeral de protección de datos, en diante, RXPd).

⁸⁸ O artigo 5 RXPd establece os principios no tratamento dos datos que poden resumirse en: (a) os datos persoais serán tratados de xeito lícito, leal e transparente en relación co interesado («licitude, lealdade e transparencia»); (b) os datos persoais só poderán ser recollidos para propósitos especificados de antemán e non poderán usarse para ningún outro propósito («limitación da finalidade»); (c) non deben recollese máis datos que os necesarios para a finalidade do tratamento («minimización de datos»); (d) debe asegurarse que os datos sexan exactos («exactitude»); (e) os datos persoais non serán almacenados máis tempo do necesario («limitación do prazo de conservación»); (f) os datos persoais deben estar protexidos («integridade e confidencialidade»).

⁸⁹ REGULAMENTO (UE) 2024/900 DO PARLAMENTO EUROPEO E DO CONSELLO do 13 de marzo de 2024 sobre transparencia e segmentación na publicidade política.

⁹⁰ Artigo 18 do Regulamento sobre transparencia e segmentación na publicidade política.

⁹¹ REGULAMENTO (UE) 2024/1689 DO PARLAMENTO EUROPEO E DO CONSELLO do 13 de xuño de 2024 polo que se establecen normas harmonizadas en materia de intelixencia artificial e polo que se modifican os regulamentos (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as directivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento de intelixencia artificial).

Estados membros) “a introdución no mercado, a posta en servizo ou a utilización dun sistema de IA que se sirva de técnicas subliminares que transcendan a conciencia dunha persoa ou de técnicas deliberadamente manipuladoras ou enganosas co obxectivo ou o efecto de alterar de maneira substancial o comportamento dunha persoa ou dun colectivo de persoas, diminuindo de maneira apreciable a súa capacidade para tomar unha decisión informada e facendo que adopten unha decisión que doutro xeito non terían tomado, dun modo que provoque, ou sexa razoablemente probable que provoque, prexuízos considerables a esa persoa, a outra persoa ou a un colectivo de persoas”, o que parece que deixa fóra da legalidade o márketing cognitivo.

Ademais, cualifícanse como de “alto risco” os sistemas de IA que elaboren perfís de persoas físicas e os sistemas de IA destinados a ser utilizados para influír no resultado dunha elección ou referendo ou no comportamento electoral de persoas físicas que exerzan o seu dereito de voto en eleccións ou referendos.

Sen dúbida, toda esta recente regulación está orientada a limitar os riscos dos usos ilícitos das ferramentas de intelixencia artificial e é moito máis do que se fixo noutros ámbitos lexislativos, pero haberá que ver se é suficiente ou se os intereses nos resultados que poden obterse infrinxindo a normativa compensan o risco.

Por outro lado, os sistemas e ferramentas de intelixencia artificial poden ser utilizados para levar a cabo campañas de desinformación masiva que, aínda que non estean centradas de xeito directo no ámbito electoral, si afectarán á formación da opinión pública. Xa sexa partindo de contidos fabricados por humanos ou fabricados por máquinas, a intelixencia artificial facilita a difusión absolutamente viral do contido apoiada por contas automatizadas que poden mesmo contestar comentarios de xeito indistinguible a como o faría un humano e adaptando o contido da mensaxe na medida necesaria para causar efecto na persoa destinataria.

Esta facilidade na difusión vese favorecida por un escenario en que os medios de comunicación perderon o seu papel como intermediarios da información, sendo substituídos polas redes sociais, onde circula todo tipo de información publicada por infinidade de fontes, tamén polos propios medios de comunicación, o que lle dificulta ao cidadán de a pé separar a información de calidade da mera desinformación.

A iso súmase o feito de que non se trata só de que o número de informadores crecese exponencialmente, mesturándose medios de comunicación con *pseudomedios*, *influencers* e opinadores en xeral, nin sequera de que o labor de contrastación sexa mellor, peor ou directamente inexistente, é que, ademais, as ferramentas de *deep learning* permiten fabricar as probas audiovisuais que se necesiten para apoiar a información que se quere difundir, como ocorreu no caso eslovaco.

A solución que ofrece o Regulamento de intelixencia artificial da Unión Europea para este asunto pasa por aumentar a transparencia, obrigando os provedores que xeren contido sintético de audio, imaxe, vídeo ou texto a marcar a información de saída do sistema de IA nun formato lexible por máquina para que sexa posible detectar que foi xerada ou manipulada de xeito artificial⁹².

⁹² Artigo 50, apartado 2 do REGULAMENTO (UE) 2024/1689 DO PARLAMENTO EUROPEO E DO CONSELLO do 13 de xuño de 2024 polo que se establecen normas harmonizadas en materia de intelixencia artificial e polo que se modifican os regu-

É dicir, trátase de incorporar como metadatos ou mediante datos cifrados nunha capa invisible ao ollo humano a información de que se trata de contido xerado por unha IA. Pero isto vai impedir que o contido se replique como veraz por centos ou miles de persoas? Certamente non, pero será posible probar que é falso, polo menos ata que alguén cos coñecementos necesarios elimine a marca de auga.

Malia todo o anterior, resulta importante ter en conta que, con todo o disruptiva que é a intelixencia artificial no momento actual, está nos albores da súa existencia; o que nos espera supera amplamente o que vimos ata o momento. Así, desde a doutrina científica sinaláronse tres grandes etapas ou tipos de intelixencia artificial que representan distintos niveis de autonomía e habilidade nas máquinas⁹³:

- 1) Intelixencia artificial estreita (ANI, pola súa sigla en inglés): refírese a sistemas de IA deseñados e adestrados para realizar tarefas específicas e limitadas. Estas intelixencias non posúen conciencia, comprensión ou verdadeiras capacidades cognoscitivas humanas. Operan dentro dun conxunto restrinxido de parámetros e non poden xeneralizar a súa aprendizaxe a dominios fóra da súa programación orixinal.
- 2) Intelixencia artificial xeral (AGI, pola súa sigla en inglés): describe unha IA capaz de entender, aprender e aplicar a súa intelixencia para resolver calquera problema de forma similar a como o faría un ser humano. Unha AGI tería habilidades cognitivas xerais, permitíndolle transferir coñecementos e habilidades entre diferentes dominios e tarefas. Ata a data, a maioría da literatura científica coincide en que a AGI non foi lograda, aínda que con certo debate⁹⁴.
- 3) Intelixencia artificial superintelixente (ASI, pola súa sigla en inglés): refírese a unha hipotética IA que supera a intelixencia humana en todos os aspectos: creatividade, sabedoría, resolución de problemas e habilidades sociais, entre outros. A ASI non só podería realizar todas as tarefas intelectuais que un ser humano pode facer, senón que o faría de xeito máis eficiente e efectivo. Este nivel de intelixencia artificial suscita importantes consideracións éticas e de seguridade, xa que o seu comportamento e decisións poderían estar máis alá da comprensión humana.

Sen dúbida, as posibilidades son practicamente infinitas e o desenvolvemento desta tecnoloxía ten o potencial de resolver os grandes problemas da humanidade, desde o cambio climático á escaseza de recursos. Cando se desenvolva a superintelixencia

lamentos (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as directivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento de intelixencia 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.

⁹⁴ A este respecto, resulta interesante a análise de Sébastien Bubeck e outros sobre a nova versión da aplicación Chat GPT, que os leva a afirmar que GPT-4 mostra indicios de intelixencia artificial xeral nas súas capacidades mentais fundamentais (como o razoamento, a creatividade e a dedución), a variedade de temas en que adquiriu experiencia (como a literatura, a medicina e a programación) e a diversidade de tarefas que é capaz de realizar (por exemplo, xogar, usar ferramentas, explicarse a si mesmo...). BUBECK, S. e outros, *Sparks of Artificial General Intelligence. Early experiments with GPT-4*. Arxiv, 2023, <https://doi.org/10.48550/arXiv.2303.12712>.

artificial, estaremos en condicións de resolver calquera problema que nos xurda; a cuestión é que unha tecnoloxía así de potente pode resultar perigosa mesmo aínda que se utilice para un obxectivo inocuo ou incluso beneficioso *a priori* para a humanidade e o planeta.

Sobre iso teorizou o filósofo Nick Bostrom, empregando a hipótese dunha IA cuxo obxectivo sexa maximizar a produción de clips. No desenvolvemento da súa tarefa, esta IA iría mellorando os procesos de produción, pero podería suceder que, buscando fabricar cada vez máis e máis clips, comezase a transformar toda a materia da Terra e despois porcións cada vez máis grandes do espazo en plantas de produción de clips⁹⁵.

De igual xeito, utilizar un sistema de IA para conseguir un obxectivo en principio lexítimo como, por exemplo, deseñar ou mellorar unha campaña electoral ou mellorar a opinión dun candidato/a nas enquisas, podería derivar doadamente nunha máquina que, en busca de cumprir o seu obxectivo, non dubidase en utilizar todos os medios á súa disposición para logralo.

Isto podería implicar non só a elaboración de perfís e deseño de propaganda electoral ultrapersonalizada, senón tamén a fabricación de calquera contido que fose necesario para apoiar o argumentario ou destruír os opoñentes, a desinformación sobre o propio proceso electoral para lograr a abstención involuntaria de certos sectores da poboación ou, mesmo, a modificación de antecedentes penais, expedientes académicos ou historiais médicos, todo canto fose necesario para lograr o seu obxectivo⁹⁶.

A principal abordaxe a este problema pasa actualmente pola aliñación (*alignment*, en inglés), tratar de aliñar os intereses e obxectivos da IA cos intereses e obxectivos dos seus deseñadores en particular e cos dos seres humanos en xeral⁹⁷ e establecer as barreiras necesarias para evitar que o sistema de IA tome decisións perigosas ou indesexables.

O problema é que, a medida que os sistemas de IA se fan máis potentes e a tecnoloxía vai evolucionando cara á intelixencia artificial xeral e á superintelixencia artificial, máis difícil resultará establecer de xeito efectivo ese aliñamento, posto que a máquina estará en condicións de prever e inventar moitos máis escenarios dos que poidan prever os seres humanos que a programen.

Faise evidente que as ferramentas coas que contamos non son suficientes para enfrontarnos aos retos que veñen. Nun escenario político convulso como o que vivimos, con conflitos armados abertos en múltiples puntos do planeta e con intereses políticos radicalmente enfrontados, a desinformación e, máis concretamente, a desinformación facendo uso de ferramentas de intelixencia artificial, configúrase

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

⁹⁶ Máis sobre a vulnerabilidade do mundo ante os avances da IA en BOSTROM, N., "The Vulnerable World Hypothesis" *Global Policy*, vol. 10:4, 2019, pp. 455-476.

⁹⁷ Este problema foi formulado en 1960 por Norbert Wiener do seguinte modo: "Se utilizamos, para alcanzar os nosos obxectivos, un medio mecánico no funcionamento do cal non podemos interferir de xeito efectivo unha vez que o iniciamos, porque a acción é tan rápida e irrevogable que non temos os datos para intervir antes de que a acción estea completa, entón é mellor que esteamos completamente seguros de que o propósito que introducimos na máquina é o propósito que realmente desexamos e non simplemente unha imitación colorida deste". 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.

como unha clara ameaza para a libre formación da opinión pública e, con iso, para o exercicio libre do dereito de sufraxio⁹⁸ e o procedemento electoral no seu conxunto, podendo supor un golpe na liña de flotación dos sistemas democráticos tal como os coñecemos.

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. 119-151.
<https://doi.org/10.5944/rdp.109.2020.29056>.
- ALONSO BETANZOS, A., “La inteligencia artificial y la prospectiva”, *Revista Atlántica de Economía*, vol. 6, n.º 1, 2023.
- ANDERSON, B. e 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. e 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. e 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 outubro de 2024).
- BOSTROM, N., “The Vulnerable World Hypothesis”, *Global Policy*, vol. 10:4, 2019, pp. 455-476.
- BUBECK, S. e outros, *Sparks of Artificial General Intelligence. Early experiments with GPT-4*. Arxiv, 2023, <https://doi.org/10.48550/arXiv.2303.12712>

⁹⁸ PAUNER CHULVI, P., 2018, *op. cit.*, p. 299: “A desinformación afecta directamente aos fundamentos das sociedades democráticas que se constrúen sobre a presunción de que os seus cidadáns toman decisións informadas no momento de acudir a votar”.

- 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. e 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 outubro de 2023 [consulta 15 de outubro 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: sobreexpuestos 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 outubro 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>
- 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. e 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>
- 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. e 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. e 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/>

regap



ESTUDIOS

- 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).
- ISAAC, J. e 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. e 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
- 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, setembro. <https://doi.org/10.4995/aniav.2022.17309>
- LEVI, S., *#FakeYou: Fake news y desinformación*, Rayo Verde Editorial, 2020.
- LINARES RODRÍGUEZ, V. e 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. e 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), 2-24, 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. e 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 e Tecnoloxía do Congreso dos Deputados (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. e 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>
- 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 setembro de 2023, actualizado 6 de marzo de 2024. [Consulta 15 de outubro de 2024]. Disponível em <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 outubro de 2023 [consultado o 15 de outubro de 2024]. Disponível em <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. e 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).

Regap



ESTUDIOS

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 (A suposta gravación da chamada entre o presidente do PS e a xornalista de Denník N mostra, segundo os expertos, múltiples signos de manipulación). AFP FACTY. 28 de setembro de 2023 [consulta 15 de outubro de 2024] Dispoñible 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

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. e 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.