Capítulo 4

La respuesta normativa

La IA avanza y se renueva a un ritmo frenético. Cualquier intento de cerrar la cuestión, dada esta condición, sería un error. El estudio del tema, a pesar de la abundancia de publicaciones, está en sus inicios, y nos encontramos ante los primeros ciclos en los que las elecciones se desarrollarán bajo la influencia de la IA. El progreso tecnológico se produce a un ritmo exponencial, trayendo consigo nuevos —e imprevisibles— desafíos para las instituciones electorales, por lo que las máximas definitivas, en este marco, tienden a quedar desfasadas e imprecisas (Archegas; Maia, 2022). De hecho, los juicios categóricos pueden sugerir lo contrario de lo que aquí se pretende: contribuir a una discusión que, por su naturaleza, debe ser abierta e inacabada. De ahí que este capítulo final no busque tanto una conclusión, sino una mirada al futuro para ver, desde el presente, los horizontes y oportunidades que el porvenir nos depara, especialmente en el ámbito normativo.

Ángel Gómez explica que la humanidad define sus etapas en función de las tecnologías más significativas: «dominar la piedra, el bronce, el hierro, el vapor, la combustión interna o los ordenadores cambió nuestra forma de vivir». Sin embargo, en todos los casos la historia ha continuado siendo «la historia de la humanidad». La IA y la integración de capas digitales en nuestras vidas, en cambio, no sólo cambian la forma de hacer las cosas, sino, «por

suerte o por desgracia», el número de cosas que nos tocan (Gómez de Ágreda, 2023, p. 222). Parte de la sustitución del hombre por la máquina, y como hemos visto, afecta al funcionamiento de las elecciones. Automatizada, la contienda por el poder político decide el destino de los hombres, desafiando en cierta medida su papel central.

La comunicación es uno de los procesos más afectados por esta transformación: la generación de contenidos sintéticos difumina la distinción entre realidad y ficción; la distribución de informaciones y opiniones depende de algoritmos inescrutables; los sistemas inteligentes buscan intereses y hábitos para dirigir mensajes perfectamente adaptados al gusto del cliente (Lewandowsky; Smillie, 2020). Todo converge para que la verdad, y la dimensión racional de la política se disuelvan en un modo de persuasión agresivo, obstinado y perturbador. La deshumanización se produce en un doble sentido: menos humanidad en el proceso de producción y en la fibra ideológica de los contenidos que circulan.

Hace años, las limitaciones técnicas y los costes incorporados dificultaban la «gestión industrial» de estos procesos, reduciendo su impacto y permitiendo el control de daños: la mecanización de la política, aunque dificultaba el mantenimiento de una comunicación (relativamente) transparente y justa, revelaba un escenario manejable, aunque desafiante. Sin embargo, la automatización y, sobre todo, la *algoritmización de la esfera dialógica* mediante la integración de soluciones de IA están reacondicionando la esfera pública, reseteando la propia dinámica de interacción entre los sujetos políticos. La IA, con sus posibilidades de aprendizaje y procesamiento del lenguaje, permite la industrialización de la persuasión, haciendo accesible —desde una perspectiva técnica y económica— la escucha masiva de las redes sociales para descubrir y falsificar códigos y formas de *hackear* las mentes.

En este contexto, cabe preguntarse si la configuración social dictada por la *plataformización de la vida* es compatible con el Estado de Derecho, la democracia y, en particular, los derechos y libertades fundamentales, dado que estas nuevas tecnologías tienden a «desestructurar los mecanismos sociales y colectivos que hemos venido organizando hasta ahora». Si bien es cierto que el

curso del siglo estará determinado por la tecnología, «también lo es que estas transformaciones deberán tener en cuenta la condición humana» y los pilares de su organización, incluido el propio sistema jurídico. Por este razonamiento, es urgente evaluar «si el Derecho se verá afectado o sustituido por otro tipo de código» en el futuro (Rebollo Delgado, 2023, p. 29).

Atentas a estas nuevas situaciones, algunas instituciones están empezando a abordar las repercusiones de la IA en la democracia. Sin embargo, como sostiene Bender, aunque los académicos han advertido de los «daños algorítmicos» en diversos ámbitos —desde el sistema de justicia penal al mercado laboral, desde la competencia en la industria al ámbito de los derechos civiles—, muchos de estos daños han arraigado en los procesos electorales, sin que hasta la fecha se hayan debatido con la profundidad, el cálculo y la atención necesarios (Bender, 2022, p. 489).

4.1. Las primeras respuestas

En sus «Meditaciones sobre la técnica», Ortega y Gasset anticipó algunas reflexiones importantes y actuales sobre el tema de la IA: para el filósofo, el ser humano no puede entenderse sin la tecnología y las posibilidades que las nuevas técnicas ofrecen al desarrollo de las civilizaciones. Está claro que las nuevas herramientas suelen traer consigo riesgos y desafíos para las sociedades. La historia está llena de ejemplos de nuevas tecnologías que, si bien aportan progreso y desarrollo, tienen el potencial de generar daños y destrucción (Rivero Ortega, 2023, p. 9), ejemplos como el fuego, el TNT, los combustibles fósiles y la energía nuclear ilustran muy bien la cuestión.

Aun así, es obvio que el recuerdo de experiencias catastróficas vinculadas a innovaciones tecnológicas no impide que la sociedad quiera explorar nuevas posibilidades y aprovechar, en la medida de lo posible, las oportunidades de ampliar la «caja de herramientas» capaz de reforzar la supervivencia en la Tierra (Rivero Ortega, 2023, p. 9). El reto reside en la difícil tarea de encontrar una coexistencia pacífica entre la libertad individual para beneficiarse

de estas nuevas tecnologías, por un lado, y el deber del Estado de proteger a la sociedad de los riesgos que entrañan estas innovaciones, por otro.

Cuando se trata de avances digitales, las intervenciones estatales a nivel normativo suelen llegar tarde (Kim, 2022, p. 52), resultando a menudo tímidas e imprecisas frente a toda la complejidad y urgencia que presenta la acelerada dinámica del entorno digital. Hay, por supuesto, algunas razones que explican este comportamiento.

La primera es que cuanto más compleja es la novedad, más difícil resulta la tarea de prever los riesgos inherentes. A pesar de la vasta literatura y de los diversos estudios en curso, la vocación vertiginosa (con la constante superposición o renovación de formas y medios de intervención) y la esencia ultraespecializada impiden una comprensión exacta de los riesgos que la IA puede generar para la humanidad a medio o largo plazo, incluso en lo que se refiere al desarrollo adecuado de los procesos electorales. De hecho, es un hecho que hoy en día:

> «[…] nos movemos en un contexto de permanente cambio. No ya cada año, sino cada mes, surge una nueva posibilidad, una mejora de aplicabilidad, se repara en una circunstancia que no había sido prevista. Las posibilidades técnicas tienen un crecimiento exponencial en cuanto a su creación y aplicabilidad en el tiempo. Por el contrario, el Derecho, como ya conocemos, es lento en su respuesta, y requiere de la actuación coordinada de muchos operadores (legislador, jurisprudencia, Administración Pública, órganos de control, etc.), a lo que se suma la necesidad de actuar también en varios niveles, como el regional o el internacional. Por muy previsores que sean los ordenamientos jurídicos, la IA tiene un fuerte componente de imprevisibilidad, de resultados o consecuencias colaterales no previstas[1], y que, en su

[1] «El aumento de su capacidad de autoaprendizaje y su grado de autonomía llevan aparejadas una relativa imprevisibilidad en su relación e interacción con su entorno, contexto y personas, que debería ser escasa o nula si realmente sometemos la IA a la pretendida seguridad, supervisión y control humano du-

> propia creación y aplicación, requieren de un proceso de control y readaptación permanente. Esto supone una dificultad añadida al Derecho como respuesta, y se constituye en el riesgo más sustantivo de la IA» (Rebollo Delgado, 2023, p. 54).

Otro factor que va en detrimento de que un movimiento global para regular la IA gane tracción es el hecho de que regular implica, en esencia, limitar en cierto sentido la libertad de explotación económica en el ámbito de tales innovaciones. Las propuestas éticas y de principios destinadas a evitar abusos o frenar excesos, por muy relevantes que sean, pueden no ser bien recibidas cuando se presentan como una forma de oponerse al progreso tecnológico global[2].

Por otro lado, es importante darse cuenta de que la tarea reguladora no es sencilla, ya que requiere conocimientos técnicos adecuados para evitar dos riesgos contrapuestos: la pasividad y el exceso regulatorio. Y es que la ausencia de un marco disciplinario que ponga coto a los abusos en la escalada digital puede ser tan perjudicial para el desarrollo social como un exagerado comportamiento garantista, fundado en una lógica paralizante, basada en una concepción equivocada de la necesaria precaución (Rivero Ortega, 2023, p. 11). Resumiendo:

> «Recordemos que el Derecho parte con unas considerables desventajas intrínsecas para regular la tecnología. Es eminentemente estatal, y hemos manifestado que las innovaciones tecnológicas

---

rante todo su ciclo de vida y, en cualquier caso, susceptible de anulación y reversibilidad de sus decisiones, acciones u omisiones cuando ello sea posible en atención al contexto de éstas, conforme recogen la mayoría de los marcos éticos que están siendo actualmente objeto de un pretendido consenso a nivel europeo e internacional […]» (Muñoz Vela, 2020, p. 26).

[2] Desde esta perspectiva, se argumenta que la restricción no es un camino viable y que sería más adecuado invertir en «mecanismos de transparencia sobre el uso de la IA por parte de los candidatos en sus comunicaciones, como la introducción de marcadores digitales». Dicho esto, hay que evitar el alarmismo y, más concretamente, «los enfoques sesgados que identifican la tecnología como [sinónimo de] fraude» (Maranhão, 2024).

> han roto la barrera del espacio y el tiempo. Otra característica del Derecho es que surge para solventar un conflicto ya existente, su capacidad predictiva es escasa y con un alto contenido de errores o portillos jurídicos, y en todo caso, normalmente se genera *a posteriori*, una vez constatada la necesidad. Por último, su elaboración está mediatizada por una infinidad de actores e intereses, a lo que se suma su dificultad de aplicación» (Rebollo Delgado, 2023, p. 16).

A pesar de las cuestiones planteadas, la democracia actual exige una ciudadanía consciente y capaz de decidir libremente y, como tal, la IA puede ser una herramienta útil para que partidos y candidatos cumplan eficazmente sus funciones informativas (Vázquez-Barrio, 2023, p. 21). Por esta misma razón, las propuestas de regulación que se han ido creando han partido de la premisa de que la IA en el ámbito electoral no puede —ni debe— ser excluida, sino que necesita de una serie de contornos jurídicos que aseguren una utilización compatible con los valores y principios que conforman la noción de integridad electoral, tales como la libertad, la igualdad, el civismo, la transparencia, la tolerancia, la sujeción a controles y el respeto a las reglas y *al juego limpio*.

Como señala Hernández Ramos (2023, p. 216) «(h)asta el momento, no hay regulaciones jurídicas omnicomprensivas de este tipo de tecnología, primando los compromisos éticos y los códigos privados de autorregulación». En términos más generales, el entorno académico comparte la opinión de que, para hacer frente a los dilemas actuales, la experiencia jurídica debe reforzarse con ajustes realizados a nivel supralegal, a través de medidas tecnológicas que, entre otras cosas, garanticen una mayor transparencia en relación con el uso de datos por parte de las plataformas digitales[3], así como protocolos más estrictos para el registro de cuen-

[3] Dolores Montero considera que el uso incontrolado de la IA plantea problemas muy graves en el ámbito del control del comportamiento. Por ello, defiende que la transparencia algorítmica debe considerarse una necesidad basada en el principio democrático, sobre todo como «respuesta al posible abuso de esta tecnología». En su opinión, la rendición de cuentas de las deci-

tas y perfiles en las redes sociales —con la vista puesta en la disipación del anonimato y el horizonte de la rendición de cuentas. En términos de la ley, los esfuerzos de transformación deben perseguir disposiciones más estrictas contra la desinformación, junto con mecanismos capaces de aumentar la responsabilidad de los actores relevantes, a la luz del potencial negativo derivado de la presencia cada vez más omnipresente de la IA (Tasioulas, 2019, pp. 87-88).

En consonancia con este espíritu, el marco regulatorio recientemente promovido por el Tribunal Superior Electoral de Brasil parece señalar que, desde una perspectiva democrática, no engañar ya debe ser visto como un deber fundamental (Garrigues Walker; González de la Garza, 2020, p. 153 y ss.; Tasioulas, 2019, p. 87), y que la protección de la verdad debe ser vista como un valor rector de la actividad legislativa (Alvim; Zilio; Carvalho, 2023, p. 80). Como tal, se torna «fundamental adoptar una estructura regulatoria que establezca el reparto de responsabilidades entre todos los actores del ecosistema digital», estableciendo «parámetros y garantías para minimizar los riesgos de la inteligencia artificial», partiendo de la premisa ineludible de que, en el ámbito electoral, es necesario establecer «acciones rápidas para prevenir ataques que pongan en riesgo el proceso electoral» (Bioni; Almeida; Mendes, 2024).

Dicho esto, a efectos didácticos, la IA puede entenderse desde dos concepciones diferentes: la *IA como (simple) herramienta* y la *IA como sujeto de la* propia comunicación. Desde esta perspectiva, Shoai y López explican que la inteligencia sintética puede actuar tanto como vehículo en el proceso de comunicación mediada (para generar, modificar o aumentar el alcance) de los mensajes publicados por sus usuarios como, alternativamente, para introducir y

---

siones tomadas por las entidades tecnológicas «es crucial para que los ciudadanos puedan, realmente, constatar que las decisiones que inciden directamente en su vida son susceptibles de ser revisadas y, llegado el caso, retiradas o modificadas». Entiende la autora que, «para ello, es fundamental que la transparencia, uno de los pilares del gobierno abierto, se manifieste también en los entornos controlados por la inteligencia artificial» (Montero Caro, 2023, p. 192).

operar agentes comunicativos que interactúen directamente con el público, a través de *bots, chatbots*[4]*, robocalls*[5] etc. (Shoai; López Molina, 2023, p. 249).

Los documentos normativos conocidos hasta la fecha abarcan, por regla general, aspectos relacionados con ambos enfoques, lo que queda ilustrado por la escasa normativa que regula, en diversos lugares, cuestiones relacionadas con la difusión de contenidos falsos o nocivos generados o distribuidos con la participación de soluciones de inteligencia sintética. En esta dirección, a la prohibición del uso de IA para producir mensajes materialmente engañosos, confirmada en Brasil y actualmente en discusión en el parlamento de los Estados Unidos (Klobuchar, 2023), se suma la prohibición de llamadas automatizadas (también en los Estados Unidos) y varias otras medidas relacionadas (Conferencia Nacional de Legislaturas Estatales, 2024).

Y es que, hasta la fecha, los proyectos regulatorios se han centrado en gran medida en la capa visible de la IA, descuidando, por regla general, cuestiones no menos importantes relacionadas con el funcionamiento discreto, silencioso y despiadado de la «selección algorítmica de mensajes» (Gerlitz; Helmond, 2013). En este

[4] Según Hampton, «la tecnología utilizada para crear *chatbots* tiene el potencial de explotar las debilidades de la arquitectura de la comunicación y obstruir los procesos políticos» (Hampton, 2019, p. 12). «La IA puede desempeñar un papel en esto filtrando y regulando la información y en forma de *bots*, que influyen en la comunicación política y potencialmente también en las preferencias de los votantes» (Coeckelbergh, 2023, p. 121). Un ejemplo es Hellovote, un *chatbot* creado en Estados Unidos para ayudar a la gente a registrarse para votar, así como para proporcionar información sobre la fecha y la ubicación de los colegios electorales (Glaser 2016). Estos *chatbots* pueden utilizarse para obtener información sobre los votantes o verificar la información que facilitan (Hampton, *ibid*.).

[5] Antes de que fuera prohibida por la FFC, esta práctica se utilizaba en Estados Unidos desde que la candidata demócrata al Congreso por Pensilvania, Shamaine Daniel, inauguró el *bot* Ashley (desarrollado por Civox), capaz de realizar llamadas automáticas gracias a la IAG. La tecnología permite realizar miles de llamadas personales al día, adaptando cada diálogo a las características e intereses del destinatario. Además, la funcionalidad puede funcionar eficazmente en más de 20 idiomas (Tong; Coster, 2023).

entorno, el pluralismo político acaba constreñido por algoritmos burbuja, y la igualdad de oportunidades se ve amenazada por una falta de transparencia que, según Óscar Sánchez, se extiende a numerosos aspectos, desde la identidad de quienes pagan hasta la cantidad de dinero gastado, desde la forma en que se filtra la información hasta la febril manera en que circulan las narrativas de desinformación (Sánchez Muñoz, 2020, pp. 87-88), en muchos casos con la ayuda de cuentas automatizadas de origen anónimo. Además, la compartición sistemática de datos personales sobre preferencias, hábitos y opiniones revela «el riesgo de manipulación de esta información, que [a menudo] se transforma en poder en manos de quienes lo detentan», en la estela de un «nuevo ciclo […] que desafía los fundamentos de la democracia» (Lins, 2023, p. 289). En efecto:

> «Si tenemos en cuenta los contextos en los que se utiliza la IA hoy en día, […] las cuestiones de privacidad y protección de datos se vuelven cada vez más problemáticas. Es relativamente fácil respetar estos valores y derechos cuando se investiga como científico social: se puede informar a los entrevistados y pedir explícitamente su consentimiento, y está relativamente claro qué ocurrirá con los datos. Pero el entorno en el que se utilizan hoy en día la IA y la ciencia de datos suele ser muy diferente. Pensemos en las redes sociales: aunque la información sobre privacidad está disponible y las aplicaciones piden a los usuarios su consentimiento, no está claro qué ocurre con los datos proporcionados ni siquiera cuáles de ellos se almacenan. Además, para utilizar la aplicación y disfrutar de sus ventajas, no hay más remedio que dar el consentimiento. A menudo, los usuarios ni siquiera se dan cuenta de que la IA está impulsando la aplicación que utilizan. A menudo, los datos proporcionados en un contexto se trasladan a otro dominio y se utilizan para un fin diferente […].
>
> Este último fenómeno también apunta al riesgo de que los usuarios sean manipulados y explotados. La IA se utiliza para manipular lo que compramos, las noticias que seguimos, las opiniones en las que confiamos, etc. Los investigadores de la teoría crítica

> han señalado el contexto capitalista en el que tiene lugar el uso de los medios sociales. Se puede decir que los usuarios de las redes sociales son "mano de obra" digital gratuita que produce datos para las empresas. […] El peligro aquí es que, incluso en las democracias actuales, la IA puede conducir a nuevas formas de manipulación, vigilancia y totalitarismo, no necesariamente como políticas autoritarias, sino de una manera más oculta y altamente eficaz: alterando la economía de tal manera que nos convierta a todos en ganado de teléfonos inteligentes ordeñados por nuestros datos»[6] (Coeckelbergh, 2023, pp. 94-95).

En este marco, a pesar de los importantes y significativos avances, sigue existiendo un significativo vacío normativo que debe ser cubierto sin más demora para que las contiendas electorales se desarrollen en condiciones más justas y equilibradas. Como señala Óscar Sánchez, en una reflexión extensible a la IA, la comunicación informatizada suscita inquietud entre los ciudadanos y conlleva un «riesgo tangible» de socavar la confianza en los procesos electorales, en la medida en que algunas de las prácticas que posibilita atacan «principios básicos sobre los que se asienta su integridad: la igualdad de armas entre los competidores electorales y la libertad de los electores para formar sus decisiones a través de un proceso de comunicación público, libre y abierto, exento de influencias indebidas». En este contexto, la falta de un marco moderno, adaptado y capaz de garantizar la plena vigencia de

[6] «Pero la IA también puede utilizarse para manipular la política de forma más directa, por ejemplo analizando los datos de las redes sociales para ayudar a las campañas políticas (como en el caso de *Cambridge Analytica*, que utilizó los datos de los usuarios de Facebook sin su consentimiento con fines políticos en las elecciones estadounidenses de 2016), o haciendo que *bots* publiquen mensajes políticos en las redes sociales basándose en el análisis de los datos de las personas en cuanto a sus preferencias políticas para influir en el voto. A algunos también les preocupa que la IA, al asumir las tareas cognitivas de los humanos, infantilice a sus usuarios, "haciéndoles menos capaces de pensar por sí mismos o de decidir por sí mismos qué hacer"» (Coeckelbergh, 2023, pp. 95-96).

estos principios ante los nuevos retos, «es un hecho sobre el que existe un consenso general» (Sánchez Muñoz, 2020, p. 21)[7].

La organización democrática, en este panorama, necesita un marco jurídico sólido y completo que tenga en cuenta el momento electoral y sus diferentes fases y adopte posiciones claras y asertivas sobre: a) la protección reforzada de la privacidad; b) la mitigación del anonimato[8]; c) el nivel de transparencia exigido (con medidas como evitar que *bots* y cuentas falsas se presenten como personas[9], identificando los contenidos generados por máquinas);

---

[7] «Con meridiana claridad lo manifiesta, por ejemplo, el estudio del Consejo de Europa sobre el uso de Internet en las campañas electorales donde se constata la "incapacidad de la regulación para garantizar un terreno de juego equilibrado para la competición política y para limitar el papel del dinero en las elecciones". Por ello, desde diferentes ámbitos institucionales y académicos se están proponiendo reformas legales de distinto carácter y alcance con el fin de embridar estas prácticas y de preservar el carácter abierto e igualitario del proceso de comunicación previo a la decisión electoral» (Sánchez Muñoz, 2020, pp. 21-22).

[8] Hay que tener en cuenta que, en muchas ocasiones, el carácter anónimo de los usuarios de las redes sociales o de los emisores de mensajes en las plataformas de *chat* es un factor de peso para fomentar el conflicto y la tensión en las manifestaciones presentes en la esfera pública, ya que la irresponsabilidad elimina lazos y barreras. Dicho esto: «El peligro que se nos aparece pronto en este contexto es que este se puede convertir en un campo adecuado para que se admitan y expandan las *fake news*, para que prolifere el "discurso de odio", o para que triunfe la "posverdad", al menos en un ámbito determinado de usuarios. Esto ocurre especialmente cuando los destinarios reciben noticias siempre con el mismo sesgo (en el entorno del llamado "filtro-burbuja"), lo que contribuye a hacer pensar al usuario que esa es la opinión más veraz, o incluso la opinión socialmente mayoritaria» (González-Torre, 2020, pp. 70-71).

[9] En Brasil, el Proyecto de Ley n.º 2.630/2020, que busca promulgar la ley de «Libertad, Responsabilidad y Transparencia en Internet», tiende a determinar que las plataformas de medios sociales y servicios de mensajería eliminen de sus ecosistemas cuentas falsas creadas con el propósito de simular la identidad de terceros y engañar al público, con la única excepción de los casos de contenido humorístico o paródico, así como las cuentas con nombre social o seudónimo. En este contexto, «las plataformas también deben prohibir las cuentas automatizadas (gestionadas por *bots*) que no se identifiquen como tales ante los usuarios. Los servicios deben aplicar medidas para identificar las cuentas que tengan movimientos incompatibles con la capacidad humana y deben

d) la responsabilidad por los contenidos producidos con IA; e) los modelos publicitarios en los que la IA juega un papel clave, para evitar que la tecnología afecte a la libertad de decisión, aumentando la transparencia (con librerías de anuncios políticos de acceso abierto) y posicionándose claramente sobre el uso del *microtargeting*, estableciendo limitaciones a su uso o incluso proponiendo la prohibición del uso de estas técnicas en procesos electorales (partiendo de la premisa de que las Constituciones no protegen el discurso de las máquinas); f) el uso de esta tecnología para combatir la desinformación (para garantizar que sea realmente independiente), así como el respeto a las garantías procesales a la hora de dejar en manos de la IA las medidas de moderación y eliminación de contenidos; y g) la responsabilidad de las plataformas, facilitando el acceso a investigadores, iniciativas de *fact-checking* y organizaciones de la sociedad civil para evaluar el impacto de la IA en las campañas políticas *online*.

Independientemente de las particularidades de los acuerdos específicos, en un sentido macro la regulación debe priorizar la integridad electoral, y no la IA *per se*. A partir de esta visión, la tarea principal será garantizar, por un lado, la eliminación de la manipulación, el abuso y el engaño y, por otro, el respeto a la libertad de expresión (garantizando que cualquier restricción respete los principios de legalidad, finalidad legítima y necesidad), la privacidad de las personas y la libertad de voto, el derecho a participar en los asuntos públicos de manera informada y en igualdad de condiciones. En consecuencia, debe evitarse la aparición de una «democracia artificial, forjada, creada […] como resultado de cuentas automatizadas y debates maximizados por la acción de robots» (Leal; Moraes Filho, 2019, p. 354), al tiempo que garantizar un cierto control sobre el funcionamiento de los algoritmos, reconociendo el hecho de que su opacidad, más allá de toda duda,

---

adoptar políticas de uso que limiten el número de cuentas controladas por un mismo usuario. El proyecto también establece que, en caso de denuncias de violaciones de la ley, uso de *bots* o cuentas falsas, las empresas podrán exigir a los responsables que confirmen su identificación, incluso mediante documento de identidad» (Lins, 2023, p. 298).

socava «la transparencia y la *rendición de cuentas* del proceso electoral», «limitando la capacidad de los individuos» (Lins, 2023, p. 303) e inclinando la balanza a favor de determinadas fuerzas o ideologías, catapultadas voluntaria o accidentalmente por la arquitectura de los medios sociales.

4.1.1. Las primeras normas

Hasta la fecha, el mosaico normativo se ha limitado fundamentalmente a instrumentos de autorregulación y *soft law* —que se han revelado insuficientes, especialmente en lo que respecta a la democracia, el Estado de Derecho y los derechos humanos (Consejo de Europa, 2020, §§ 18 a 20)—, sin embargo, como hemos visto, diversas instituciones y organizaciones internacionales han trabajado para establecer marcos de protección democrática de alcance general, adaptados a este nuevo contexto.

Se trata de un buen planteamiento, ya que la respuesta normativa al uso de la IA en los procesos electorales no puede provenir exclusivamente de la regulación electoral, ni puede relegarse —sin reservas— a la autorregulación de los proveedores de tecnología[10]. En consecuencia, el inminente tratamiento de la cuestión en el ámbito particular del Derecho Electoral no dispensa del diseño de soluciones generales a los retos de la IA, especialmente en un plano normativo que vaya más allá de las fronteras nacionales.

En definitiva, se considera necesario un marco normativo general, con principios, reglas de contenido y una estructura institucional capaz de abordar el impacto de la IA no sólo en los derechos fundamentales, sino también en la propia organización política y social (Montilla Martos, 2023, p. 167). También hay que tener en cuenta que abordar los dilemas del contexto digital supone colisiones con campos adyacentes al mundo del derecho, como la

[10] Al fin y al cabo, según el orden natural de las sociedades modernas, «que sean las nuevas tecnologías las que se adapten al dogma tradicional del Derecho y no el dogma jurídico el que sea adaptado a las exigencias de las nuevas tecnologías» (Coello de Portugal, 2014, p. 70).

ingeniería de *software* y la gobernanza interna de las plataformas digitales[11].

Idealmente, los proyectos con este alcance deberían abordar, entre otras, cuestiones: la gobernanza de los datos, el derecho al honor, la intimidad y la propia imagen (*deepfakes*), la cuestión del anonimato, el discurso del odio, el acoso y la violencia política, la protección de los grupos vulnerables, la mitigación de los riesgos sistémicos y la rendición de cuentas transfronteriza, así como un apartado específico sobre la comunicación electoral, con normas eficaces para garantizar la transparencia, la rendición de cuentas y protecciones reforzadas en el ámbito de la extracción de información personal y la individualización de contenidos influyentes (*microtargeting*), junto con otras medidas para preservar la igualdad entre competidores. El objetivo primordial es, como sabemos, la neutralidad de las tecnologías y, para garantizarla, hay que invertir en apertura, controlabilidad y explicabilidad, así como garantizar la participación humana en el diseño, la fijación de objetivos, la aplicación y la supervisión de toda la cadena de actividades que implican a la IA.

[11] «Parece claro que la nueva sociedad digital no puede ser ahormada únicamente por el Derecho, teniendo en cuenta que supone una nueva conformación horizontal del conjunto social, y que afecta a los medios de comunicación, a la economía, a todos los ámbitos del social, incluso a las mismas bases de organización social y política. Por ello parece muy acertada la propuesta de De la Quadra Salcedo, cuando usa el concepto de la solución holística y de forma concreta manifiesta que "Esa afección a todos los elementos fundamentales que estructuran e informan nuestras sociedades hace obligatorio adoptar una perspectiva holística en el tratamiento de los retos que plantea la sociedad digital". Atribuir al Derecho la única forma de ordenar y encajar la sociedad digital es un grave error. El derecho debe ser, como ha sido siempre, una forma de solventar conflictos sociales con una perspectiva de bien común, pero en todos casos necesita de la colaboración de otras áreas de conocimiento, de todos los elementos que conforman la estructura social» (Rebollo Delgado, 2023, p. 52).

La respuesta de la Unión Europea

En la Unión Europea no existía hasta hace poco tiempo una regulación específica sobre campañas electorales. Sin embargo, a estas les es de aplicación el RGPD, que afecta al «tratamiento de datos personales en el contexto de unas elecciones» estableciendo una serie de previsiones y limitaciones como el consentimiento explícito, el tratamiento de datos para cumplir una obligación legal o de interés público, etc., incluidas las obligaciones de los responsables del tratamiento de datos. Según establece el RGPD, los actores políticos «están obligados a informar a las personas sobre el tratamiento de sus datos y muy especialmente sobre la identidad del responsable, los fines del tratamiento, fuentes de los datos, destinatarios, etc. Obsérvese que en el contexto de unas elecciones la serie de datos recogidos y tratados pertenecen muchos de ellos a la categoría de sensibles, por lo que tanto los encargados como los responsables del tratamiento tienen que aplicar medidas adecuadas para garantizar el nivel de seguridad que se exige en proporción con los riesgos que entrañan» (García Mahamut, 2023).

También se aplica la Directiva 2022/2555 del Parlamento europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión, que al recalcar que las entidades pertenecientes al sector de las infraestructuras digitales se basan esencialmente en sistemas de redes y de información, subraya la necesidad de la seguridad física de dichos sistemas como parte de sus medidas para la gestión de los riesgos de ciberseguridad ya que, como hemos visto, su utilización en período electoral puede atacar la integridad del proceso y provocar desconfianza en el funcionamiento del mismo.

Además, la reciente aprobación de un reglamento sobre IA, que establece normas armonizadas sobre IA y modifica las normas comunitarias anteriores, también afecta a las campañas electorales, y ha venido acompañada de la adopción de otras medidas específicas para este momento determinante de la democracia. Este reglamento, como señala Simón Castellano (2023, p. 26) «establece un modelo de regulación basada en la responsabilidad

proactiva de los distintos actores que participan en el desarrollo, implementación y comercialización de las herramientas basadas en IA, y lo hace con un enfoque que parte de la delimitación de niveles de riesgo en función de la tecnología empleada y de sus usos posibles».

Para hacer frente a estos riesgos específicos, el Reglamento Europeo de Inteligencia Artificial prevé cuatro modalidades[12], cada una de ellas con una directriz que se describe a continuación:

> «*Riesgos mínimos o inexistentes*: la mayoría de los sistemas de IA con riesgos insignificantes pueden seguir funcionando sin regulación;
>
> *Riesgos limitados*: los sistemas de IA con riesgos aceptables están sujetos a ligeras obligaciones de transparencia para que los usuarios puedan tomar decisiones con conocimiento de causa;

[12] «Para profundizar en la regulación basada en el riesgo, es necesario retroceder unos pasos para explorar la noción […]. Gellert (2017) conceptualiza el riesgo como una herramienta que ayuda al proceso de toma de decisiones, dirigiendo su análisis no a su existencia —presunta—, sino a cuánto riesgo puede o está dispuesto a asumir un determinado agente y cuánto es capaz de mitigar. Del mismo modo, Hood *et al.* (2001) definen el riesgo como una probabilidad de consecuencias adversas, siendo la regulación del riesgo una interferencia gubernamental en el mercado o en los procesos sociales para controlar las posibles consecuencias adversas. […] Hood *et al.* señalan que la actividad humana y la tecnología en los tiempos modernos tienen el efecto colateral de riesgos que dependen de expertos para ser evaluados y reconocidos, son colectivos, globales e irreversibles en su impacto, dando lugar a una "sociedad del riesgo", distinta de períodos históricos anteriores. Por tanto, el riesgo puede entenderse como "la capacidad de definir lo que puede ocurrir en el futuro y elegir entre alternativas", funcionando como herramienta para la toma de decisiones en la medida en que hace cierto lo incierto. Sus elementos constitutivos son dos operaciones distintas pero vinculadas: predecir el futuro (con ayuda de números) y tomar decisiones en función de ello. Así, el riesgo, aunque asociado a algo más cuantificable, también puede entenderse como un elemento cualitativo y valorativo que necesita ser evaluado desde diferentes perspectivas» (Bioni; Garrote; Guedes, 2023, p. 25).

> *Riesgos elevados*: se autorizará un amplio espectro de sistemas de IA de alto riesgo, pero con requisitos y obligaciones estrictos para acceder al mercado de la UE[13];
> *Riesgos inaceptables*: se prohibirán, con limitadas excepciones, los sistemas que contengan riesgos considerados inaceptables, como la manipulación cognitiva, la vigilancia policial predictiva, el reconocimiento de emociones en lugares de trabajo y escuelas, la puntuación social y determinados sistemas de identificación biométrica a distancia» (Parlamento Europeo, 2023; Revoredo, 2023; Ryan-Mosley, 2024)[14].

En lo que se refiere al ámbito electoral, en primer lugar, en el nuevo Reglamento se prohíben los *riesgos inaceptables,* así como los que amenazan la seguridad o los derechos y libertades de los ciudadanos, incluidas las prerrogativas relacionadas con el ejercicio libre, consciente y autodeterminado de la participación política. En esta categoría se incluyen las aplicaciones capaces de manipular el comportamiento humano[15] e identificar o proporcionar información sobre las vulnerabilidades de determinados grupos, así como las circunstancias especiales que implican la categorización

---

[13] Los sistemas capaces de afectar negativamente a la seguridad o a los derechos fundamentales se consideran de alto riesgo. Entre ellos se incluyen: a) los sistemas de IA sujetos a la normativa de seguridad de productos (automóviles, sistemas de aviación, dispositivos médicos, ascensores, etc.); y b) los sistemas de IA relacionados con la gestión del funcionamiento de infraestructuras críticas, la gestión de la inmigración, el acceso a servicios públicos esenciales, la gestión de prestaciones estatales, etc.

[14] En cuanto a los plazos de cumplimiento, se espera que las normas prohibitivas entren en vigor a finales de 2024, mientras que las normas que imponen obligaciones a las empresas que desarrollan «modelos fundacionales» (modelos que sirven de base para otros productos de IA, como GPT-4) tendrán que cumplir la ley en el plazo de un año. Las demás obligaciones derivadas de la nueva legislación deberán cumplirse en un plazo de dos años (Ryan-Mosley, 2024).

[15] Ejemplos de estas prácticas son las técnicas de marketing y la publicidad subliminal, que se introducen en la conciencia de las personas para alterar sustancialmente su comportamiento de forma potencialmente perjudicial para sus intereses (Sotero, 2023, p. 162).

biométrica o la videovigilancia masiva por parte de las autoridades en espacios públicos.

También afectan a los procesos electorales los *sistemas de alto riesgo*, que amenazan infraestructuras críticas y pueden interferir en los derechos de las personas, como los relacionados con los derechos democráticos (gestión del censo electoral, reconocimiento de firmas en el voto por correo o sistemas de identificación biométrica para el acceso al voto). Todos los sistemas que utilicen estas técnicas deberán garantizar: a) la gobernanza de los datos, de forma que mantengan estándares de calidad y estén libres de sesgos y discriminaciones; b) la seguridad y la supervisión humana en todos los ciclos[16]; c) el cumplimiento de sus deberes de transparencia sobre el funcionamiento del sistema; d) el registro en una base de datos a nivel comunitario; y e) la superación del test de conformidad, con vistas a la obtención de la correspondiente certificación.

Por último, los sistemas de riesgo medio o bajo, como los asistentes virtuales o *chatbots* que no afectan directamente a la privacidad, inicialmente no suponen riesgos significativos para los derechos y libertades, aunque eventualmente puedan ser utilizados en aplicaciones de recomendación de voto (por cierto, cada vez más habituales). En esta dimensión, la medida básica garantiza la transparencia, para que los usuarios puedan entender cómo funcionan y sus principales características y puedan evitar sesgos ideológicos o partidistas.

Además, recientemente, y con motivo de las elecciones al Parlamento Europeo de 2024, se han adoptado una serie de medidas que, aunque no tratan específicamente de la IA, contienen referencias concretas a sus posibles aplicaciones en las campañas electorales. Las medidas se derivan del «Plan de Acción Europeo para la Democracia» (2020), que busca «empoderar a los ciudadanos y

[16] Según Fernanda Lage, «el ciclo de vida de la IA incluía técnicamente las siguientes fases: 1. Diseño, datos y modelización (planificación, recopilación de datos y construcción del modelo); 2. Desarrollo y validación (entrenamiento y pruebas); 3. Despliegue; 4. Supervisión y perfeccionamiento (solución de cualquier problema que se produzca)» (Lage, 2022, p. 62).

aumentar la resiliencia democrática en toda la Unión promoviendo elecciones libres y justas, reforzando la libertad de los medios de comunicación y combatiendo la desinformación». El Plan se desarrolló por fases, con hitos en noviembre de 2021 y diciembre de 2023, y contó con la cooperación del Parlamento Europeo, especialmente en la lucha contra la injerencia extranjera y la desinformación, a través del trabajo de sucesivas Comisiones Especiales sobre el tema. Entre las medidas observadas, cabe destacar las siguientes:

a) El Código Reforzado de Buenas Prácticas contra la Desinformación, adoptado en 2022 (Comisión Europea, 2022), que sustituye y refuerza el Código anterior (Comisión Europea, 2018) y fue firmado por 34 entidades, incluidas plataformas en línea, agencias de publicidad, verificadores de hechos, instituciones académicas y organizaciones de la sociedad civil. Sus compromisos, que serían de aplicación al uso de la IA, incluyen: desmonetizar la difusión de la desinformación; garantizar la transparencia de la publicidad política; reducir el comportamiento no auténtico utilizado para difundir la desinformación; cooperar con los verificadores de hechos; y proporcionar a los investigadores acceso a los datos. Dentro del grupo de trabajo creado para supervisar su aplicación, destaca la creación de un Centro de Transparencia[17], que recoge informes de las plataformas que forman parte de él[18].
b) El Reglamento sobre transparencia y segmentación de la publicidad política[19] busca apoyar unas elecciones libres y justas y también afecta a actividades para las que se está utilizando ya IA. De esta manera: a) amplia el concepto de

[17] Disponible en: [https://disinfocode.eu], consultado el: 11-04-2024.

[18] Disponible en: [https://disinfocode.eu/reports-archive/?years=2024], consultado el: 11-04-2024.

[19] Disponible en: [https://data.consilium.europa.eu/doc/document/PE-90-2023-INIT/es/pdf], consultado el: 11-04-2024. Para un estudio detallado del texto (García Mahamut, 2023).

publicidad política; b) reduce la fragmentación jurídica y elimina los obstáculos habituales de los servicios transfronterizos; c) aumenta las obligaciones de transparencia de la publicidad política (definiendo como tal la que realizan los partidos políticos pero también los anuncios temáticos), que deberá identificarse como tal y ofrecer información básica sobre el patrocinador, las elecciones a las que está vinculada, el importe invertido y las técnicas de segmentación utilizadas y d) restringe el uso de técnicas de microsegmentación y amplificación para este tipo de publicidad política, que a partir de ahora sólo podrá realizarse con datos recabados directamente del sujeto, con su consentimiento explícito y distinto para este uso. También, e) prohíbe, en cualquier caso, la microsegmentación basada en datos personales que afecten a la raza, la etnia o las opiniones políticas. Y, por último, en un esfuerzo por evitar injerencias externas, f) prohíbe la contratación de publicidad por parte de organizaciones de terceros países durante los tres meses anteriores a la votación, en línea con «la Sentencia del Tribunal General de 25 de noviembre de 2020 en el asunto T-107/19 en la que se afirma que un partido de un país que no pertenece a la UE no entra en la definición de "partido político" al no tratarse de ciudadanos de la UE ni estar reconocido por el ordenamiento jurídico de al menos de un Estado miembro, o establecido de conformidad con este» y el Dictamen del Tribunal de Cuentas Europeo 01/2022 sobre la propuesta de la Comisión de Reglamento sobre el estatuto y la financiación de los partidos políticos europeos y las fundaciones políticas europeas (García Mahamut, 2023; p. 90).

c) La Ley de Servicios Digitales (DSA) también tiene un efecto directo en las elecciones, ya que regula la moderación de los contenidos en línea y armoniza las normativas nacionales sobre contenidos ilegales, publicidad y desinformación, algo habitual, como hemos visto, en campaña electoral. En particular: a) establece mecanismos que permiten a los usuarios denunciar este tipo de contenidos; b) garantiza que las

decisiones tomadas por los moderadores de las plataformas puedan ser impugnadas y c) refuerza la transparencia de las plataformas, incluyendo, por ejemplo, la transparencia de los algoritmos utilizados para las recomendaciones. Para las plataformas muy grandes, que llegan a más del 10% de la población de la UE[20], se define explícitamente la obligación de mitigar los riesgos sistémicos relacionados con los procesos electorales, lo que ha dado lugar a una serie de recomendaciones específicas para las próximas elecciones al Parlamento Europeo, entre ellas: e) reforzar sus procesos internos en función del riesgo potencial; f) mejorar su capacidad para dar respuesta a estos comportamientos; g) implementar medidas de mitigación de riesgos; h) promover la información oficial sobre procesos electorales, por ejemplo, con iniciativas de alfabetización mediática y i) adaptar sus sistemas de recomendación para empoderar a los usuarios y reducir la monetización y viralidad de contenidos que amenacen la integridad de los procesos electorales. Además, según el nuevo Reglamento sobre transparencia y orientación de la publicidad política, comentado anteriormente, la publicidad política debe estar claramente etiquetada como tal, registrando explícitamente cuándo se produce el uso de IAG. También se fomenta la cooperación con las autoridades nacionales y de la UE, expertos independientes y organizaciones de la sociedad civil, especialmente el Grupo de Trabajo del Observatorio Europeo de Medios Digitales (EDMO). Por último, se recomienda adoptar un mecanismo para responder a los incidentes que puedan tener un impacto significativo en los resultados o en la participación electoral y evaluar la eficacia de las medidas mediante análisis postelectorales, fomentando la supervisión por terceros para garantizar que las medidas aplicadas son eficaces y respetan los derechos fundamentales.

[20] Disponible en: [https://digital-strategy.ec.europa.eu/en/policies/list-designated-vlops-and-vloses], consultado el: 11-04-2024.

d) Por último, la Recomendación sobre Procesos Electorales Inclusivos y Resilientes en la UE aborda la protección y la ciberseguridad de las infraestructuras relacionadas con las elecciones, que deben clasificarse como críticas, así como las bases de datos y los procesos, en línea con la propuesta de Ley de Resiliencia Cibernética. Propone medidas para minimizar los riesgos de injerencia de terceros países mediante el control de la financiación de los partidos políticos y sus campañas, y refuerza la transparencia de partidos y candidatos. Entre las medidas incluidas están: la adopción de códigos de conducta que faciliten la integridad electoral y campañas justas, promoviendo un discurso político inclusivo y prohibiendo comportamientos manipuladores, como la generación o difusión de falsedades (incluyendo *deepfakes*), o que inciten al odio, así como el rechazo al uso de tácticas, técnicas y procedimientos no auténticos para difundir o amplificar mensajes políticos, donde la IA juega un papel cada vez más importante; promoviendo mecanismos de control independientes para el cumplimiento de los compromisos adquiridos. Por último, se anima a los Estados a proteger el entorno informativo y garantizar que los votantes reciban información correcta, promoviendo proyectos de sensibilización y alfabetización mediática para hacer frente a la manipulación informativa, las injerencias y la desinformación relacionadas con las elecciones, así como reforzando las respuestas rápidas, tanto *prebunking* como *debunking*.

En resumen, las próximas elecciones europeas tendrán lugar en un contexto de guerra y la UE ha creado un marco normativo que pretende evitar la desinformación y la injerencia extranjera y, dentro de su propuesta integral, trata de dar respuesta a la utilización de la IA en las campañas electorales. Todo el mundo tendrá la oportunidad de observar la eficacia de un marco jurídico —representado por un ambicioso conjunto de medidas— que, aunque ha llegado un poco tarde y no podrá aplicarse plenamente hasta 2025, formará parte del marco normativo-protector de las eleccio-

nes que se celebren a partir de entonces en los Estados miembros y probablemente en muchos otros.

Norteamérica

Dentro del enfoque general, desde 2019, Canadá ha regulado la toma de decisiones automatizada[21] para reducir los riesgos relacionados con errores o discriminación. Esta regulación establece el principio de transparencia, con la obligación de publicar en un lugar destacado que la decisión será tomada por un sistema automatizado; la necesidad de hacer público cualquier código fuente en poder del Gobierno, así como una serie de medidas preventivas, como pruebas previas para detectar sesgos involuntarios en los datos, procesos de seguimiento de los resultados de estos sistemas de toma de decisiones y la garantía de intervención humana, así como el derecho a recurrir esta clase de decisiones.

En Estados Unidos, además de las distintas regulaciones en los distintos Estados[22], a finales de 2023, se publicó la *Orden Ejecutiva sobre el Desarrollo y Uso Seguro y Confiable de la Inteligencia Artificial,* que obliga a los principales desarrolladores de IA y proveedores de computación en la nube a compartir pruebas e información relacionada con asuntos de seguridad nacional, así como otras cuestiones consideradas cruciales para los asuntos gubernamentales.

---

[21] Directiva sobre la toma de decisiones automatizada de 1 de abril de 2019 (disponible en: [https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592], consultado el: 4-5-2024).

[22] National Conference of States Legislatures (2022). Legislation Related to Artificial Intelligence (disponible en [https://www.ncsl.org/technology-and-communication/legislation-related-to-artificial-intelligence3], consultado el: 4-4-2024); U.S. Chamber of Commerce (2022). State-by-State Artificial Intelligence Legislation Tracker. Interactive map shows states' action to legislate artificial intelligence (disponible en [https://www.uschamber.com/technology/state-by-state-artificial-intelligence-legislation-tracker?state=4], consultado el: 4-4-2024).

Concebido con el objetivo de conciliar los aspectos positivos y negativos derivados de la IA ante el «rápido avance de sus capacidades», el decreto busca «hacer frente a los desafíos tecnológicos y contribuir a la prosperidad y seguridad de las personas», considerando que el «uso irresponsable» de la tecnología en boga «podría exacerbar daños sociales como el fraude, el trato discriminatorio y la desinformación», imponiendo amenazas «a la libre competencia, el mercado laboral y la seguridad nacional» (Estados Unidos, 2024). Según el decreto, la política de desarrollo y uso responsable implica que la utilización de las herramientas de IA debe:

> 1. *Garantizar la protección y la seguridad*, lo que requiere evaluaciones, políticas, instituciones y mecanismos normalizados, sólidos, repetibles y fiables para mitigar los riesgos antes de su uso. Estos requisitos recaen principalmente en las aplicaciones de la IA en campos como la biotecnología, la ciberseguridad, las infraestructuras críticas (como el suministro energético) y otros asuntos que afectan a la seguridad nacional. Para ello, los ensayos, las evaluaciones y las pruebas de rendimiento posteriores al lanzamiento pueden proporcionar una base sólida para abordar los riesgos de la IA sin sacrificar sus beneficios.
>
> 2. Promover la innovación y la competencia responsable para que Estados Unidos sea líder en IA.
>
> 3. Apoyar a los trabajadores estadounidenses mediante la creación de nuevos empleos e industrias, e incorporar la negociación colectiva para garantizar que la clase trabajadora se beneficie de estas oportunidades. En este sentido, el decreto menciona la provisión de formación laboral para apoyar la oferta de mano de obra cualificada. También pretende garantizar que la aparición de la IA no vulnere los derechos ni empeore la calidad del trabajo, ni fomente una vigilancia indebida de los trabajadores ni introduzca nuevos riesgos para su salud o seguridad.
>
> 4. Explorar cómo la IA podría discriminar ilegalmente o facilitar la administración de programas y prestaciones estatales (a nivel federal) para promover la equidad y los derechos civiles.

> 5. Proteger los intereses de los consumidores aplicando la legislación vigente y promulgando nuevas salvaguardias contra el fraude, los daños involuntarios, la discriminación, la violación de la intimidad y otros posibles perjuicios derivados de la IA.
>
> 6. Proteger la privacidad y las libertades civiles garantizando que la recogida, uso y almacenamiento de datos se realiza de forma legal y segura, mitigando los riesgos para la privacidad y la confidencialidad.
>
> 7. Gestionar los riesgos derivados del uso de la IA por parte del gobierno federal, reforzando su capacidad interna para regular, gobernar y apoyar el uso responsable de la IA.
>
> 8. Liderar el progreso social, económico y tecnológico mundial, incluido el liderazgo efectivo en sistemas pioneros y salvaguardias para el despliegue responsable de la IA. Esta acción incluye comprometerse con aliados y socios globales para desarrollar un marco que mitigue los riesgos de la IA, libere su potencial positivo y se una para superar los retos compartidos» (Dougall; Ostrowski, 2024).

En lo que se refiere específicamente al uso de IA en las elecciones, la FCC en Estados Unidos, tras el uso de los audios falsos del Presidente Biden en *robocalls* durante las primarias de New Hampshire, ha declarado ilegales estos *robocalls* que utilizan voces generadas por IAG[23]. Además, en el Congreso de los Estados Unidos se han presentado, al menos cuatro proyectos de ley que abordan específicamente el uso de *deepfakes* y otros contenidos manipulados en las elecciones federales y al menos otros cuatro que abordan estos contenidos de forma más amplia. A nivel estatal, en los últimos años se han aprobado nuevas leyes que prohíben o restringen de otro modo los *deepfakes* y otros medios engañosos en la publicidad electoral y los mensajes políticos en estados tan

[23] Disponible en: [https://docs.fcc.gov/public/attachments/FCC-24-17A1.pdf], consultado el: 4-4-2024.

ideológicamente diversos como California[24], Minnesota[25], Texas[26], Washington[27] y Florida[28].

California, tras asegurarse de que los procedimientos relacionados con las elecciones tengan prioridad en la agenda judicial, prohíbe la distribución de medios audiovisuales engañosos que afecten a un candidato dentro de los 60 días previos a una elección, a menos que se incluya una advertencia indicando que el medio ha sido manipulado, y establece procedimientos legales para los candidatos cuyas imágenes o voces sean utilizadas de manera engañosa. Aunque, proporciona exenciones para ciertos medios de comunicación.

Minnesota y Texas establecen el delito de usar tecnología de *deep fake* para influir en una elección, dentro de los 90 días anteriores a la misma en Minnesota y de los 30 días en Texas. En ambos es necesaria la intención de dañar a un candidato o influir en el resultado de una elección, y se establecen diferentes penas según la gravedad del delito.

Washington, quizás en la regulación más completa, además de exigir a los tribunales una respuesta rápida, facilita al perjudicado acceder a medidas cautelares u otro tipo de alivio equitativo para prohibir la publicación de dichos medios sintéticos, presentar una acción por daños generales o especiales contra el patrocinador. Establece la obligación de etiquetar el contenido. Por último, establece la responsabilidad de los patrocinadores de comunicaciones electorales que contienen medios sintéticos, pero no de los medios

---

[24] Disponible en: [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730], consultado el: 4-4-2024.

[25] Disponible en: [https://www.revisor.mn.gov/bills/text.php?number=HF1370&type=bill&version=3&session=ls93&session_year=2023&session_number=0], consultado el: 4-4-2024.

[26] Disponible en: [https://capitol.texas.gov/tlodocs/86R/billtext/html/SB00751F.htm], consultado el: 4-4-2024.

[27] Disponible en: [https://lawfilesext.leg.wa.gov/biennium/2023-24/Pdf/Bills/Session%20Laws/Senate/5152-S.SL.pdf?q=20231003125542], consultado el: 4-4-2024.

[28] Disponible en: [https://www.flsenate.gov/Session/Bill/2024/919/BillText/er/PDF], consultado el: 4-4-2024.

que los difunden, excepto en ciertas circunstancias y exime a los proveedores o usuarios de servicios de computación interactiva de ser tratados como el editor o portavoz de cualquier información proporcionada por otro proveedor de contenido de información, pero permite que sean responsables en ciertas circunstancias.

Por último, Florida se limita a establecer la obligación de etiquetado de cualquier pieza que haya utilizado IAG «en todo o en parte», exigiendo que además de la utilización de estas técnicas se genere la falsa apariencia de una persona real, y se pretenda atacar a un candidato o «engañar sobre una cuestión electoral», estableciendo con gran detalle las condiciones en las que debe aparecer este aviso en distintos formatos, como texto escrito, televisión o video, internet, audio, o cualquier comunicación gráfica. Cualquier persona identificada como responsable de un anuncio político que no incluya el descargo de responsabilidad requerido comete un delito menor de primer grado, sancionable por ley.

4.1.2. Autorregulación

Aunque «(p)or norma general, los compromisos éticos se limitan a suscribir una autodeclaración, que describe una estructura interna con un comité de ética y un consejo asesor, pero su función y los poderes siguen sin estar claros» (Hernández Ramos, 2023, p. 216), en un contexto de vacío normativo, en el ámbito electoral en los últimos meses se han lanzado dos iniciativas por parte de diferentes empresas tecnológicas relacionadas con el uso de la tecnología en las elecciones: a) el *Acuerdo Tecnológico para Combatir el Uso Engañoso de la Inteligencia Artificial en las Elecciones de 2024*[29]; y b) las *Directrices Voluntarias sobre Integridad*

[29] Anunciado en febrero de 2024, el acuerdo en cuestión fue firmado por Adobe, Amazon, Anthropic, ARM, ElevenLabs, Google, IBM, Inflection AI, LinkedIn, McAfee, Meta, Microsoft, Nota, OpenAI, Snap, Stability AI, TikTok, TrendMicro, TruePic y X. Se puede acceder a su texto completo en la siguiente dirección: [https://www.aielectionsaccord.com/].

Electoral para Empresas Tecnológicas, estipuladas por la *Fundación Internacional para Sistemas Electorales* (IFES)[30].

El Acuerdo Tecnológico incluye una serie de acciones y compromisos para neutralizar la difusión de contenidos producidos con sistemas de IA que puedan «comprometer la integridad de los procesos electorales». Materiales como imágenes, audio y vídeo generados con herramientas basadas en IA «que falsifican o alteran de forma engañosa la apariencia, la voz o las acciones de candidatos políticos, funcionarios electorales y otras partes interesadas». Los compromisos son:

> «1. Desarrollar y aplicar tecnología para mitigar los riesgos relacionados con contenidos electorales engañosos creados con sistemas de IA, incluidos los de código abierto.
> 2. Evaluar los modelos de IA en el ámbito del presente acuerdo para comprender los riesgos que pueden plantear en relación con la producción de contenidos electorales engañosos.
> 3. Controlar y detectar la distribución de tales materiales en sus plataformas.
> 4. Tomar medidas para tratar adecuadamente la información engañosa distribuida en sus servicios (etiquetado).
> 5. Fomentar la adaptación entre los sectores sensibles a los contenidos electorales engañosos.
> 6. Proporcionar al público información que haga transparentes las medidas de mitigación.
> 7. Colaborar con organizaciones académicas y de la sociedad civil para desarrollar planes de seguimiento.

[30] Las directrices se lanzaron en marzo de 2024 y han sido respaldadas por los organismos electorales de Australia, Mauritania, Haití, Filipinas, República Dominicana, Rumanía y El Salvador, así como por expertos de diferentes países, organizaciones internacionales como la OEA, organizaciones de la sociedad civil como Freedom House, el Atlantic Council, The Carter Center o el German Marshall Fund e IDEA. Además, algunas empresas tecnológicas como Google, Meta, Microsoft, Snap y TikTok han hecho suyos los principios establecidos. El documento completo puede consultarse en la siguiente dirección: [https://electionsandtech.org].

8. Apoyar los esfuerzos para promover la concienciación pública, la alfabetización mediática y la resiliencia en toda la sociedad».

Las Directrices voluntarias, por su parte, establecen un marco para: a) mejorar la relación entre las autoridades electorales y las empresas tecnológicas; b) establecer políticas y procesos claros; c) compartir información; d) garantizar que los votantes tengan acceso a información de alta calidad; e) mejorar los canales de comunicación entre las empresas tecnológicas y las autoridades electorales; f) democratizar, facilitando que empresas de todo tipo y tamaño contribuyan a la integridad electoral. Los compromisos son:

«1. Determinar cómo dar prioridad a la participación en elecciones en todo el mundo mediante un proceso que tenga en cuenta una serie de factores, incluidos los derechos humanos y los principios democráticos, el uso pertinente de los productos y servicios de la empresa y consideraciones relativas a los recursos.

2. Consultar con la sociedad civil mundial, cuando sea necesario y apropiado, a través de los canales o eventos establecidos, con el fin de informar a las empresas sobre la comprensión del contexto electoral nacional y el compromiso con las autoridades electorales, según proceda.

3. Establecer y dar a conocer políticas y procesos claros sobre el contenido, las actividades, los disturbios civiles y la violencia relacionados con las elecciones.

En la medida de lo posible, haga que estas políticas y procesos sean accesibles de forma adecuadamente localizada.

4. Centralizar la información sobre políticas, productos y servicios que puedan ser útiles para las autoridades electorales y la sociedad civil.

5. Establecer procesos de planificación que tengan en cuenta los plazos y la capacidad de las autoridades electorales, incluso antes del período electoral y después de las elecciones.

Identificar los puntos de contacto bilaterales en una fase temprana, organizar la coordinación y tener en cuenta la capacidad local.

6. Permitir el acceso a información fiable sobre las elecciones y los votantes, cuando proceda.

7. Establecer una estrategia para hacer frente a la desinformación y la información errónea sobre la participación electoral.

Publicitar estas políticas y darlas a conocer a las autoridades electorales y otras partes interesadas.

8. Establecer y poner a disposición de las autoridades electorales canales de comunicación que puedan utilizarse para tratar incidentes críticos durante el período electoral.

9. Proporcionar públicamente información sobre contenidos políticos y/o electorales pagados en todo el mundo.

Proporcionar información que pueda facilitar los esfuerzos de investigación sobre cuestiones relacionadas con contenidos políticos y/o electorales pagados.

10. Mantener los mecanismos y operaciones de coordinación adecuados más allá del período inmediatamente posterior a las elecciones.

11. Apoyar el compromiso postelectoral con las partes interesadas en las elecciones.

12. Apoyar el análisis postelectoral de las autoridades electorales y otras partes interesadas en las elecciones, según proceda».

En ambos casos, los principios y compromisos son genéricos y no contemplan explícitamente, por ejemplo, la prohibición o eliminación de *las deepfakes*. Tampoco incluyen detalles sobre cómo se aplicarán las resoluciones, ni siquiera un calendario de actuación.

Por su parte, como hemos visto, Meta[31] se ha comprometido a etiquetar las imágenes creadas con IA en todas sus plataformas, además ha declarado que prohibirá que las campañas políticas utilicen sus nuevos productos de publicidad generados por IA y requerirá que los anunciantes políticos revelen cuando utilicen

[31] Disponible en: [https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/], consultado el: 4-5-2024.

herramientas de IA para modificar o crear anuncios en Facebook e Instagram. Su Oversight Board[32], organismo creado por la propia compañía para revisar las decisiones de Meta sobre moderación de contenido, en un documento frecuente de advertencias sobre el papel de la plataforma social en los procesos electorales señala la necesidad de establecer estándares claros para el contenido generado por IAG y *fakes* (*deep* o *cheap*).

Google por su parte, además de etiquetar este tipo de contenidos en YouTube[33], ha restringido, desde comienzos de 2024, las respuestas relacionadas con las elecciones en su *chatbot* Gemini, ofreciendo como respuesta «Todavía estoy aprendiendo a contestar esta cuestión. Mientras, puedes consultar Google Search», restringiendo también la experiencia de búsqueda generativa relacionada con las elecciones, aplicando estas restricciones desde principios de 2024.

Por último, OpenAI[34], creadora de ChatGPT y DALL-E, se compromete a prevenir el abuso, fomentar la transparencia y asegurar la integridad de las elecciones en todo el mundo. Tratando de anticiparse y prevenir el mal uso potencial de sus herramientas, especialmente en el contexto electoral. Para lograrlo ha puesto en marcha una serie de medidas para detectar y abordar el abuso, como identificar contenido generado por IA y prevenir la creación de *chatbots* que imitan a candidatos. Con esa intención, se introducirá un sistema de credenciales y marca de agua digital de la Coalición para la Proveniencia y Autenticidad del Contenido (C2PA) para identificar imágenes generadas por IA y, han anunciado el lanzamiento de un clasificador de procedencia para detectar imágenes generadas por DALL-E, incluso si han sido modificadas. Además, se implementará una función de «reporte»

---

[32] Disponible en: [https://www.oversightboard.com/wp-content/uploads/2024/04/Oversight-Board-Elections-Paper-May-2024FINAL.pdf], consultado el: 4-5-2024.

[33] Disponible en: [https://www.npr.org/2023/11/14/1212986395/youtube-will-label-ai-generated-videos-that-look-real], consultado el: 4-5-2024.

[34] Disponible en: [https://openai.com/blog/how-openai-is-approaching-2024-worldwide-elections], consultado el: 4-5-2024.

para usuarios, permitiéndoles señalar posibles violaciones en el uso de GPTs personalizados.

4.2. La respuesta de los organismos electorales (Brasil)

Ante la ausencia de una respuesta normativa algunos organismos electorales han empezado a tomar medidas para ofrecer una. Entre todos, por su carácter integral y reglamentario destaca Brasil, donde ante la ausencia de un marco normativo —y en un escenario de riesgos inminentes, el Tribunal Superior Electoral, aprovechando la prerrogativa inscrita en el artículo 57-J de la Ley de Elecciones (Ley n.º 9.504/97[35]), aprobó, el 27 de febrero de 2024, un conjunto de reglas destinadas a regular el uso de IA en las campañas, concretamente en los arts. 9.º-B a 9.º-H de la Resolución n.º 23.610/2019, actualizada por la Resolución n.º 23.372/2024, actualizada por la Resolución n.º 23.372/2024, dando así un paso importante en la defensa activa de la integridad de los procesos de renovación política y de la propia democracia brasileña[36]. Por

[35] «Artículo 57-J. El Tribunal Superior Electoral reglamentará lo dispuesto en los artículos 57-A a 57-I de esta Ley de acuerdo con el escenario y las herramientas tecnológicas existentes en cada momento electoral y promoverá, para los vehículos, partidos y demás entidades interesadas, la formulación y amplia divulgación de reglas de buenas prácticas relativas a las campañas electorales en Internet».

[36] Disponible en: [https://www.tse.jus.br/legislacao/compilada/res/2024/resolucao-no-23-732-de-27-de-fevereiro-de-2024], consultado el: 4-7-2024. Ante la traumática experiencia del 8 de enero, era «más que urgente pensar en regular el uso de la inteligencia artificial con fines políticos y partidistas. Si ya existía una considerable preocupación por la manipulación del flujo informativo desde la elección de Trump y Bolsonaro […], tales temores se han vuelto aún más relevantes con los crecientes avances de la inteligencia artificial, especialmente en su modalidad generativa, que posibilita incluso la llamada "propaganda sintética". Ante estos riesgos, el TSE emitió recientemente una serie de resoluciones para tratar de evitar la creación y difusión de contenidos falsos en las elecciones, además de imponer obligaciones de transparencia, para que los destinatarios de contenidos producidos con inteligencia artificial sean alertados de esta circunstancia» (Frazão, 2024).

su novedad, su carácter pionero y su relevancia en el ámbito del estudio del impacto de la IA en las campañas electorales vamos a realizar una exposición detallada de su contenido.

De acuerdo con el resumen de Ana Frazão, el texto normativo, tal como fue aprobado, establece como objetivos principales:

> «1. Prohibir *deepfakes*.
> 2. Advertir sobre el uso de la IA en la propaganda electoral.
> 3. Restringir el uso de robots para mediar en el contacto con los votantes, prohibiendo la simulación de un diálogo con un candidato o cualquier otra persona.
> 4. Responsabilizar a las *grandes tecnológicas si* no retiran inmediatamente los contenidos que contengan desinformación, discursos de odio, ideología nazi y fascista, así como contenidos antidemocráticos, racistas y homófobos»[37] (Frazão, 2024).

[37] En un resumen alternativo y centrado específicamente en la producción de piezas con IA, la resolución establece tres tipos de uso, guiados por normas diferentes: «El primer grupo incluye los usos ordinarios, por ejemplo, para realizar ajustes en imágenes y sonido o para crear viñetas y otros elementos gráficos. El uso de estas tecnologías es libre y no requiere identificación. La norma tiene sentido, ya que hoy en día cualquier fotografía tomada con un teléfono móvil, *v. gr*, pasa automáticamente por un proceso de ajuste gráfico mediante herramientas de IA, algo que no suscita mayores reacciones en el público. La creación de contenidos sintéticos/artificiales o ajustes más avanzados, en cambio, requieren que la campaña informe claramente al electorado de que el material se ha producido utilizando IA. Existe aquí un nivel intermedio de regulación, que supone que la información transmitida a los receptores del mensaje les permitirá hacer un juicio más realista de lo que se presenta. Por último, hay usos que están estrictamente prohibidos, y pueden llegar a constituir abuso de poder económico o uso indebido de los medios de comunicación, lo que puede dar lugar a la anulación de candidaturas y mandatos, así como a la imposición de la pena de inelegibilidad durante ocho años. Esto incluye las *deepfakes*, independientemente de su finalidad —ya sean positivas o destinadas a atacar a candidatos contrarios—, así como la difusión de desinformación potenciada por el uso de IA. También están prohibidos los *chatbots* y similares que pretendan hacerse pasar por los propios candidatos, dando la impresión al electorado de haber entablado una comunicación directa con esa persona» (Neisser; Mattiuzzo, 2024).

4.2.1. Autorización de uso de la IA

Para empezar, el artículo 9-B[38] autoriza expresamente el uso de la IAG para la producción de contenidos[39], así como el uso de soluciones de inteligencia sintética para la mejora, edición o adaptación de materiales de comunicación. La IAG ha sido expresamente autorizada con fines de creación (concepción completa), sustitución (modificación de componentes estéticos, sonoros o de contenido), omisión (supresión de líneas, participaciones, elementos visuales o detalles adyacentes), *blending* (mezcla de audio, unión de secuencias), alteración de la velocidad (reducción del tiempo de visualización), superposición de imágenes (por ejemplo, para que una grabación de estudio se proyecte sobre una grabación externa) o de sonidos (por ejemplo, para que la voz de un locutor prevalezca sobre el discurso de otro agente, mostrado simultáneamente).

No obstante, por regla general, la existencia de contenidos (total o parcialmente) sintéticos debe señalarse mediante una *cláusula de exención de responsabilidad*, cuya finalidad es garantizar que la audiencia no sea engañada o inducida a error, es decir, que el público en general no tenga dudas sobre el origen o la naturaleza del material informativo. En particular, el Alto Tribunal asu-

---

[38] «Artículo 9-B. La utilización en propaganda electoral, en cualquiera de sus formas, de contenidos sintéticos multimedia generados mediante inteligencia artificial para crear, sustituir, omitir, fusionar o alterar la velocidad o superponer imágenes o sonidos impone al responsable de la propaganda el deber de informar, de manera explícita, destacada y accesible, que el contenido ha sido fabricado o manipulado y la tecnología utilizada».

[39] La IAG puede definirse como «una metodología usada […] para describir cualquier tipo de inteligencia artificial que contenga algoritmos de aprendizaje no supervisado para crear nuevas imágenes digitales, vídeo, audio, texto o código. El propósito es generar *datos sintéticos* que puedan pasar una prueba de Turing […]». Dentro de esta vertiente, el salto evolutivo se produce con la introducción de los métodos de redes generativas adversariales (o redes generativas antagonistas) —GANs— conceptualizados como «un tipo de algoritmo de inteligencia artificial que funciona como un juego en el que una red neuronal compite con otra para generar imágenes o contenido nuevo» (Moreno, 2023, p. 36 y 60).

me el espíritu de la recomendación contenida en un informe elaborado por la Comisión Europea, que señala la importancia de la transparencia como forma de reducir las posibilidades de manipulación (Denemark, 2024, p. 130)[40].

Según el §1 del artículo comentado a) en los medios de audio, la advertencia debe preceder al mensaje principal (ya que los oyentes pueden no seguirlo hasta el final); b) en las imágenes estáticas (registros fotográficos, representaciones realistas, artes gráficas) la advertencia debe formar parte de la composición visual, con el uso de una etiqueta o marca de agua (considerando la posibilidad de que las personas no lean los subtítulos), además de ser reforzada, cuando se publica en medios digitales, por mecanismos de audiodescripción[41]; c) en los medios de audio y vídeo (vídeo con audio), el requisito se combina, exigiendo, acumulativamente, una advertencia al principio de la pieza, el uso de una etiqueta o marca de agua y la audiodescripción; y d) en el caso de materiales impresos (como guías, folletos de propuestas o cuadernillos), la advertencia debe reproducirse en todas las páginas (teniendo en cuenta la posibilidad de que los lectores no recorran todo el material).

---

[40] Sin embargo, en contra de las ambiciones del escenario europeo, el Tribunal Superior Electoral perdió la oportunidad de exigir la presencia de advertencias legales también en los mensajes con contenido político personalizados sobre la base de actividades de minería de datos y elaboración de perfiles psicométricos. Según los expertos, esta norma sería una de las más importantes para evitar que las técnicas de IA conduzcan a discursos nocivos y a la desinformación (Denemark, *ibidem*).

[41] La audiodescripción es un recurso tecnológico capaz de traducir imágenes en palabras y es necesario, en particular, para satisfacer las necesidades de las personas con discapacidad visual, ya sean ciegas o con baja visión. En términos técnicos, la audiodescripción puede implementarse, entre otras formas, mediante: a) fusión con el audio original, ofreciéndose como opción al oyente o espectador; b) disponibilidad en auriculares (como es habitual, por ejemplo, en museos); c) incrustada en textos ocultos a los que se puede acceder, opcionalmente, en dispositivos o *software* de lectura (como Kindle); y d) a través de *hashtags* cada vez más presentes en las redes sociales, como #*paraciegover*.

Sin embargo, la propia resolución estipula supuestos en los que se prescinde de la advertencia (artículo 2)[42], empezando por cambios estéticos neutros, dirigidos, por ejemplo, a eliminar ruidos, corregir la luz, aumentar el pixelado, etc. Se entiende, no obstante que la flexibilización en cuestión, habida cuenta de la destacada importancia de la imagen como determinante cognitivo del voto (Alvim, 2019, p. 258), no se extiende a los ajustes dirigidos a reconstruir positivamente la imagen de los candidatos, lo que puede apreciarse, por ejemplo, en las técnicas de *de-aging* (rejuvenecimiento artificial) o *makeover* (embellecimiento o aumento del capital de simpatía o atractivo).

También se renuncia a la necesidad de informar en el caso de los símbolos electorales distintivos, como logotipos o viñetas, lo que se justifica por el hecho de que estos elementos, en sentido estricto, cumplen una simple finalidad de identificación, pasando más o menos por alto las tácticas de información, persuasión, movilización o convencimiento.

Por último, las advertencias de IA no son necesarias para los montajes habituales, como los que reúnen en una sola toma a candidatos con patrocinadores o simpatizantes políticos, insertan un fondo falso sobre *croma* o alteran una foto de estudio para situar al candidato en cualquier entorno.

El apartado 3 del artículo 9b[43] permite expresamente el uso de avatares[44] o *chatbots* (aplicaciones conversacionales) en las cam-

---

[42] «El párrafo 2 de este artículo no se aplica a: I – los ajustes destinados a mejorar la calidad de la imagen o del sonido; II – la producción de elementos gráficos de identidad visual, viñetas y logotipos; III – los recursos de *marketing* que se utilizan comúnmente en las campañas, como el montaje de imágenes en las que candidatos y simpatizantes aparecen en un único registro fotográfico utilizado en la producción de material publicitario impreso y digital».

[43] «Parágrafo 3.º El uso de *chatbots*, avatares y contenidos sintéticos como artificio para mediar la comunicación de campaña con personas naturales se sujeta a lo dispuesto en el *caput* de este artículo, quedando prohibida cualquier simulación de diálogo con el candidato u otra persona real».

[44] Los avatares digitales no son más que personajes ficticios, humanoides o no, creados digitalmente con fines comunicativos, que interactúan con los usuarios a través de texto o, más frecuentemente, de alguna forma de habla, espe-

pañas brasileñas, siempre que además vayan acompañados de una advertencia efectiva a los usuarios, que elimine cualquier tipo de duda sobre el origen sintético de las interacciones establecidas[45]. En consecuencia, corresponde a los competidores que lo utilicen dejar claro a los usuarios que los mensajes y respuestas ofrecidos por el *chatbot* derivan de comandos automáticos y que, por lo tanto, no dialogan directamente con el candidato, con ninguno de sus colaboradores de campaña ni con ninguna persona física[46]. El texto normativo, en definitiva, prohíbe la simulación de interacciones humanas, pero bajo el imperativo de la transparencia autoriza el uso de esta herramienta.

Para los responsables, el incumplimiento de estos requisitos conlleva la posibilidad de retirada de contenidos o la indisponibilidad del servicio de comunicación, no sólo ante una orden judicial, sino también por propia iniciativa de las plataformas, en la modalidad del artículo 9-B, § 4, de la Resolución. Además, la sanción de retirada o desconexión puede ir acompañada de otras amonestaciones previstas en la legislación, de modo que el uso no advertido de avatares para difundir mensajes ilegales (difamatorios, calumniosos, perjudiciales, de odio o desinformativos) puede aca-

---

cialmente en medios virtuales. En el ámbito electoral, la estrategia ganó visibilidad con la presentación de iXóchitl, la vocera virtual de Xóchitl Gálvez, precandidata a la presidencia de México, en diciembre de 2023 (Urquijo, 2023).

[45] «[…] el uso de *chatbots*, avatares y contenidos sintéticos como artificio para mediar en la comunicación de campaña con personas físicas prohíbe cualquier simulación de interlocución con el candidato u otra persona real. Esto nos remite a la reciente decisión de la Comisión Federal de Comunicaciones, el organismo regulador del sector de las telecomunicaciones en Estados Unidos, que declaró ilegales las llamadas telefónicas con recursos de voz generados por Inteligencia Artificial. En un comunicado, la FCC subrayó que la tecnología de clonación de voz se ha utilizado en llamadas automatizadas para extorsionar a familiares vulnerables, hacerse pasar por famosos y desinformar a los votantes» (De Teffé, 2024).

[46] Dado que los *chatbots* pueden utilizarse —y a menudo se utilizan— para recopilar datos personales de los votantes, se deduce que ofrecer este tipo de servicios conlleva la obligación implícita de cumplir las normas aplicables a la protección de datos en el contexto electoral.

rrear la multa aplicable a los casos de anonimato, según el artículo 57-D, § 2, de la Ley Electoral.

4.2.2. Prohibiciones del uso electoral de la inteligencia artificial

El artículo 9-C de la norma prohíbe el uso, en la publicidad, de contenidos artificialmente fabricados o manipulados para difundir narrativas desinformativas capaces de desequilibrar o comprometer la integridad de la elección[47]. La norma prohíbe explícitamente la desinformación contra las instituciones electorales y, en consecuencia, dialoga con el artículo 2 de la Resolución 23.714/2022[48], que regula el ejercicio del poder político en caso de contenidos falsos o descontextualizados que socaven la confianza pública en el procedimiento en cuestión. La combinación de ambas disposiciones demuestra que: a) la desinformación antisistema puede detenerse tanto mediante órdenes administrativas basadas en el poder de policía, independientemente de la provocación, como a partir de los procesos judiciales por propaganda ilegal que tienen su origen en reclamaciones presentadas por actores habilitados para ello; y b) la desinformación contra can-

[47] «Artículo 9-C Queda prohibida la utilización, en la propaganda, cualquiera que sea su forma o modalidad, de contenidos fabricados o manipulados para difundir hechos notoriamente falsos o descontextualizados con potencial para causar daño al equilibrio de la elección o a la integridad del proceso electoral».

[48] «Artículo 2.º En los términos del Código Electoral, se prohíbe divulgar o compartir hechos notoriamente falsos o gravemente descontextualizados que afecten la integridad del proceso electoral, incluyendo los procesos de votación, escrutinio y cómputo de votos. § Párrafo 1.º Verificada la hipótesis prevista en el encabezamiento, el TSE, en decisión motivada, ordenará a las plataformas la retirada inmediata de la URL, URI o URN, bajo pena de multa de entre R$ 100.000,00 (cien mil reales) y R$ 150.000,00 (ciento cincuenta mil reales) por hora de incumplimiento, a partir del final de la segunda hora después de recibida la notificación. § Párrafo 2.º: Entre el día anterior y los tres días siguientes a la elección, la multa del Párrafo 1.º se aplicará a partir del final de la primera hora después de recibida la notificación».

didatos o partidos, en principio[49], queda fuera del ámbito del poder de policía y debe tratarse principalmente en el contexto de proceso judiciales por propaganda ilegal, basados en el artículo 96 de la Ley Electoral.

Cabe destacar que la norma prohíbe los contenidos falsos o contextualizados que generen un desequilibrio en la contienda, invocando así una expresión con un doble significado. En primer lugar, desequilibrio puede entenderse como aquello que socava las condiciones de igualdad de la elección, generando o intensificando distancias que, a la postre, condicionan las posibilidades de victoria de los distintos participantes. En este sentido, el *caput* prohíbe específicamente las *fake news* que tengan el potencial de dañar la reputación de cualquier competidor (desinformación contra candidatos y partidos). Sin embargo, el concepto de desequilibrio también puede asociarse a las ideas de agitación, falta de control o inestabilidad. Desde este punto de vista, el artículo 9 *quáter* también protege la normalidad del proceso, prohibiendo los contenidos falsos con polémica dirigidos a crear o reforzar la hostilidad intergrupal, la «violencia moralista» (Aguado Terrón; Villaplana Jiménez, 2023, p. 206) y el linchamiento digital (Vallespín, 2021, p. 78), fomentando la división, el conflicto y la animadversión social (Lardiez, 2021, p. 15) contra los órganos de supervisión y control (desinformación contra las instituciones).

---

[49] Excepcionalmente, sin embargo, la desinformación contra candidatos o partidos puede ser objeto de objeción espontánea, cuando se caracteriza la hipótesis descrita en el artículo 4 de la Res. TSE N.º 23.714/2022: «Artículo 4 La producción sistemática de desinformación, caracterizada por la publicación constante de información falsa o descontextualizada sobre el proceso electoral, autoriza la suspensión temporal de perfiles, cuentas o canales mantenidos en las redes sociales, con sujeción a los requisitos, plazos y consecuencias establecidos en el artículo 2. Párrafo único. La orden a que se refiere el encabezamiento incluirá la suspensión del registro de nuevos perfiles, cuentas o canales por los responsables o bajo su control, así como el uso de perfiles, cuentas o canales de contingencia previamente registrados, bajo pena del delito previsto del Código Electoral».

El apartado 1 del artículo 9-C[50] establece que los contenidos audiovisuales de base sintética no podrán utilizar registros biométricos (voz o imagen) de figuras humanas, vivas, fallecidas o inventadas[51], aunque, en los dos primeros casos, se obtenga autorización de la persona clonada para ello. La redacción, sin embargo, es dudosa, ya que no deja claro si la prohibición de clonar con IA es absoluta, aplicándose a todos los casos (leyendo el §1 de forma aislada) o si, en sentido contrario, sólo se aplicaría cuando el mensaje que favorece o beneficia a un candidato contiene elementos desinformativos (leyendo el §1 en relación con el titular del artículo).

La cuestión es bastante importante, ya que la primera interpretación lleva a la conclusión de que la presencia de seres humanos en vídeos sintéticos sólo está permitida en piezas de contenido neutro (por ejemplo, para la difusión de información general, como las agendas de campaña), pero no en productos persuasivos, que en definitiva constituyen los vídeos realmente valiosos para las campañas. En consecuencia, las técnicas generativas quedarían restringidas incluso a los vídeos de animación, dado que la norma analizada, leída aisladamente, prohíbe textualmente no sólo la representación sintética de personas vivas o muertas, sean famosas, anónimas o históricas, sino también el acto de «crear […] una persona […] ficticia».

La lectura alternativa, en cambio, invita a ello por dos razones: en primer lugar, porque el contexto topológico, al menos en teoría, lleva a interpretar el §1 a la luz de la norma que abre el artículo (encabezamiento); en segundo lugar, porque el §1 termina con una

---

[50] «Apartado 1 Prohíbe el uso, para perjudicar o favorecer una candidatura, de contenidos sintéticos en formato audio o vídeo, o una combinación de ambos, que hayan sido generados o manipulados digitalmente, incluso con autorización, para crear, sustituir o alterar la imagen o la voz de una persona viva, fallecida o ficticia (*deep fake*)».

[51] Cabe señalar que el uso de la IAG para generar vídeos realistas que exploren la figura de líderes políticos fallecidos ha sido habitual en todo el mundo, y ha proliferado como estrategia de campaña electoral en varios países, como la India (Mukherjee, 2024), como ha pasado con piezas de publicidad de Sergio Masa en las elecciones presidenciales de Argentina (Durães, 2023).

referencia explícita a la expresión «*deep fake*» (*deepfake*). Si, según este modo de pensar, el §1 se considera una orden que interactúa con el encabezamiento según un principio de continuidad, la conclusión sería que el contenido sintético puede contar con la presencia de seres humanos, siempre que no transmita contenido desinformativo. Sin embargo, esta conclusión sería inadecuada, sobre todo porque el contenido desinformativo ya está prohibido en sí mismo, independientemente de cualquier otro aspecto de forma o fondo (artículo 9-C, *caput*). La norma, bajo esta apariencia, sería superflua, inocua y redundante.

En consecuencia, se entiende que, por el momento, el TSE, en una posición conservadora, ha optado por prohibir la creación sintética de representaciones humanas con el ánimo de evitar que los votantes sean inducidos a error al interpretar a los personajes artificiales como simpatizantes, partidarios o portavoces de carne y hueso. En otras palabras, dentro de la comunicación persuasiva, la IAG se liberó exclusivamente para contenidos claramente ficticios, excluyendo el hiperrealismo como medida para evitar fraudes o errores autoinducidos, lo que denota un temor implícito a una etapa de superación de una versión moderna del Test de Turing, históricamente relacionado con el descubrimiento de la capacidad de las máquinas para actuar como humanos sin ser reconocidas[52].

Aunque la resolución (artículo 9-C, § 2[53]) establece que el incumplimiento de esta norma «constituye» una hipótesis de abuso

[52] Creada por Alan Turing, destacado investigador en los campos de la informática y la IA, la prueba se asemeja a un juego de imitación, con tres participantes, dos humanos y un ordenador. «El evaluador, un humano, hace preguntas abiertas a los otros dos (un humano y un ordenador) con el objetivo de determinar cuál de ellos es el humano. Si el evaluador no puede hacer la distinción, se asume que el ordenador es inteligente. […] La genialidad de este concepto es que no hay necesidad de comprobar si la máquina sabe realmente algo, si es consciente de sí misma o incluso si es correcta. En su lugar, la prueba de Turing indica que una máquina puede procesar grandes cantidades de información, interpretar el habla y comunicarse con los seres humanos [sin ser descubierta]» (Taulli, 2020, pp. 18-19).

[53] «Apartado 2: El incumplimiento de lo dispuesto en el encabezamiento y en el apartado 1 de este artículo constituye abuso de poder político y utilización

de poder o de utilización indebida de los medios de comunicación, se argumenta que la anulación del mandato o la declaración de inelegibilidad deben interpretarse como una posibilidad, y no como una imposición determinista. Esto se debe a que la fijación de una gravedad máxima por presunción absoluta sería inconstitucional porque violaría los principios de proporcionalidad, razonabilidad e individualización de las penas, además de entrar en conflicto con lo dispuesto en el artículo 22, inciso XVI, de la Ley Complementaria n.º 64/1990[54], que impone, para estos casos, la necesidad de un examen caso por caso, regido por una evaluación precisa de la «gravedad de las circunstancias»[55].

---

indebida de los medios de comunicación, dando lugar a la anulación de la inscripción o del mandato, e impone la declaración de responsabilidad en los términos del apartado 1 del artículo 323 del Código Electoral, sin perjuicio de la aplicación de otras medidas que procedan en relación con la irregularidad de la propaganda y la ilegalidad del contenido».

[54] «Artículo 22: Cualquier partido político, coalición, candidato o Ministerio Público Electoral podrá realizar una representación ante la Justicia Electoral, directamente ante el Corregidor General o Regional, denunciando hechos e indicando pruebas, indicios y circunstancias, y solicitar la apertura de una investigación judicial para investigar el uso indebido, mal uso o abuso del poder económico o del poder de autoridad, o el uso indebido de vehículos o medios de comunicación, a favor de un candidato o partido político, de acuerdo con el siguiente procedimiento: […] XVI —para la configuración del acto abusivo no se considerará la potencialidad del hecho para alterar el resultado de la elección, sino sólo la gravedad de las circunstancias que lo caracterizan».

[55] Además, la posibilidad de una infracción penal, a pesar de la mención expresa, no siempre se plantea, ya que el artículo 323 del Código Electoral tiene la propaganda «en relación con partidos o candidatos» como elemento del tipo. Por lo tanto, la desinformación contra las instituciones electorales puede dar lugar a la investigación de abuso de poder, así como ser objeto de una representación que solicite su destitución. Pero no puede ser investigada penalmente desde la perspectiva de este delito.

4.2.3. Obligaciones de las plataformas

El artículo 9-D[56] impone a las plataformas[57] la obligación de adoptar y publicar medidas para prevenir o reducir la circulación de narrativas desinformativas que puedan poner en peligro el buen desarrollo de las elecciones. Aunque no se utiliza el término, en este sentido la resolución establece un deber de cuidado[58], deter-

[56] «Artículo 9-D. Es deber del proveedor de aplicaciones de Internet, que permita la difusión de contenidos político-electorales, adoptar y publicitar medidas para evitar o reducir la circulación de hechos notoriamente falsos o gravemente descontextualizados que puedan afectar la integridad del proceso electoral, entre ellas: […]».

[57] En sentido estricto, la imposición recae sobre cualquier plataforma «que permita la difusión de contenidos políticos electorales». En la práctica, sin embargo, no hay noticias de plataformas que prohíban debates sobre estos temas, por lo que, a pesar de la señalización potencialmente restrictiva, la norma tiene aplicación general.

[58] El deber de diligencia consiste en una «obligación jurídica que establece que una persona u organización es responsable de sus acciones u omisiones que causen daños a terceros». Considerado como un «principio fundamental de la responsabilidad civil», el deber de diligencia «exige que las personas y organizaciones adopten medidas razonables para evitar daños a terceros, y se aplica a cualquiera que pueda verse suficientemente afectado por las acciones u omisiones de la persona u organización en cuestión» (Campos; Oliveira; Santos, 2023). A raíz de los debates relacionados con la positivización del deber de cuidado en diversos contextos, como el estadounidense y el europeo, desde hace algunos años las empresas tecnológicas están introduciendo cambios en la forma de presentar determinados contenidos para minimizar el impacto de la desinformación, especialmente adaptando los algoritmos que seleccionan y ordenan la información que visualizan los usuarios. En este sentido: «YouTube […] ha hecho desde finales de 2016 algoritmos que deciden los contenidos que se sugieren automáticamente después de una visualización, añadiendo un nuevo criterio de responsabilidad social. De esta manera se trata de corregir la tendencia a favorecer la recomendación de vídeos con contenidos extremos para maximizar las visualizaciones, algo de lo que se había acusado frecuentemente a esta compañía. También ha eliminado millones de canales por violación de sus directrices y ha empezado a mostrar en las primeras posiciones de las búsquedas más contenidos de fuentes autorizadas y pertenecientes a medios de comunicación tradicionales. Además, […] han puesto en marcha iniciativas para fomentar el periodismo de calidad y el *fact-checking*. […] Facebook ha tomada

minando la adopción de una postura proactiva dirigida a prevenir, o al menos actuar, ante la presencia de desinformación contra las instituciones electivas en el entorno digital, claramente con el objetivo de hacer frente a una nueva escalada de movimientos antidemocráticos, teniendo en cuenta el recuerdo reciente del traumático intento de golpe de Estado que tuvo lugar el 8 de enero de 2023. Entre otras cuestiones, estas medidas apuntan a un debate centrado en la mejora de sus normas comunitarias y, especialmente, en la dinámica de moderación de los mensajes. A este respecto, cabe señalar que:

> Todas las grandes empresas de medios sociales han creado aparatos específicos para la moderación de contenidos. La mayoría adoptan un enfoque de servicio al cliente: se encarga a los usuarios (y, cada vez más, a los programas informáticos) que identifiquen contenidos o comportamientos problemáticos; a continuación, los moderadores de la plataforma llevan a cabo una revisión procedimental, entre bastidores, y deciden si eliminan o no las publicaciones señaladas basándose en sus propias directrices y juicios. Puede decirse que estos procedimientos, construidos a lo largo de más de una década, han sido capaces de satisfacer las necesidades de las plataformas, manteniendo al mismo tiempo la promesa de una comunidad suficientemente segura para un número satisfactorio de usuarios […].
>
> Esa sería la mejor versión de la historia. Otra versión dice que una cultura tóxica de acoso, especialmente contra las mujeres y las minorías, parece haber arraigado en las redes sociales, siendo tolerada alegremente por los gestores de las plataformas, deseosos de fomentar sus propias ideas de libertad de expresión y beneficiarse de los datos que recogen por el camino; que los legisladores en Europa y otros lugares exigen intervenciones más

---

diferentes iniciativas para reducir la propagación de noticias falsas y, en general, de contenido inauténtico, adaptando los algoritmos internos para que los contenidos procedentes de sitios sospechosos tengan menos presencia en la sección de noticias (*feeds*) de los usuarios» (Sánchez Muñoz, 2020, p. 115).

> estrictas de las plataformas contra la incitación al odio y la propaganda terrorista; que una avalancha de noticias fraudulentas y teorías conspirativas erosionan la confianza e influyen en la forma de pensar de los votantes, hasta un punto tal vez suficiente para decidir elecciones nacionales. Incluso un alto nivel de éxito en la moderación a gran escala sigue permitiendo cientos de miles de errores y cientos de miles de descuidos, y cada uno de ellos representa un usuario agraviado, engañado o desprotegido» (Gillespie, 2020, pp. 329-330).

Aunque bien intencionada, la norma es controvertida, entre otros factores porque: a) establece un deber expreso al margen de un precepto legal (la prerrogativa inscrita en el artículo 57-J de la Ley Electoral debe leerse en consonancia con el artículo 105 de la misma ley)[59]; b) la potestad reglamentaria existe para desarrollar, y no para contradecir el ordenamiento vigente[60], chocando apa-

---

[59] «Artículo 105. Hasta el 5 de marzo del año de la elección, el Tribunal Superior Electoral, teniendo en cuenta el carácter reglamentario y sin restringir derechos ni establecer sanciones distintas a las previstas en esta Ley, podrá dictar todas las instrucciones necesarias para su fiel cumplimiento, oyendo previamente, en audiencia pública, a los delegados o representantes de los partidos políticos».

[60] Como explica la doctrina, el papel del TSE en el ejercicio de su función reguladora es reducido, ya que está sujeto a las siguientes limitaciones: «a) limitación de carácter material: al tener el rango de ley ordinaria, no pueden ocuparse de materias que la Constitución reserva a una ley complementaria, como las relativas a la inelegibilidad y a la organización de la Justicia Electoral; b) limitación de carácter lógico: si bien tienen fuerza de ley, su carácter reglamentario no les permite contradecir actos normativos primarios; (c) limitación de naturaleza política: prohibición de sustituir la función encomendada por la Constitución al Poder Legislativo, so pena de violar el principio republicano; y (d) limitación temporal: deben dictarse en el límite máximo del 5 de marzo del año electoral» (Alvim, 2016, p. 70). En resumen: «La Resolución puede ser entendida como una fuente formal secundaria, que se encarga de interpretar y regular las fuentes primarias, como la Constitución y las leyes federales. No puede innovar en el ordenamiento jurídico ni restringir derechos o establecer sanciones distintas de las previstas en ley. […] En este caso, la ley autoriza expresamente la reglamentación del asunto por Resolución. De acuerdo con el artículo 57-J de la Ley 9.504/97, el TSE reglamentará lo dispuesto en los artícu-

rentemente la norma con el régimen de responsabilidad de los intermediarios previsto en el artículo 19[61], *caput*, del Marco Civil de Internet[62], y reflejado en la propia legislación electoral, en el artículo 57-F[63] de la Ley Electoral[64]; c) el deber de actuar para

---

los 57-A a 57-I de esta Ley (capítulo de la propaganda electoral en Internet) según el escenario y las herramientas tecnológicas existentes en cada momento electoral y promoverá, para los vehículos, partidos y demás entidades interesadas, la formulación y amplia divulgación de reglas de buenas prácticas relativas a las campañas electorales en Internet. Sin embargo, como se ha señalado, es importante que las normas propuestas dialoguen directamente con las normas vigentes en el ordenamiento jurídico» (De Teffé, 2024).

[61] «Artículo 19. Con el fin de garantizar la libertad de expresión y evitar la censura, el proveedor de aplicaciones de Internet sólo podrá ser considerado civilmente responsable de los daños y perjuicios derivados de contenidos generados por terceros si, ante una orden judicial previa y expresa, no adopta medidas para que los contenidos señalados como infractores no estén disponibles en el ámbito y límites técnicos de su servicio y en el plazo indicado, sin perjuicio de las disposiciones legales en contrario».

[62] Como consecuencia de esta nueva norma, la plataforma de aplicaciones de Internet sólo podrá ser considerada responsable en el ámbito civil «si, tras una orden judicial específica, no adopta las medidas necesarias para que el contenido identificado como nocivo no esté disponible dentro del ámbito y los límites técnicos de su servicio y en el plazo especificado. Esta regla resuelve, al menos en parte, la divergencia entre sentencias judiciales que condenan o no a los proveedores de servicios de Internet por el contenido de páginas ofensivas en sus sitios web y redes sociales. La finalidad es garantizar la plena libertad de expresión en el uso de Internet, impidiendo cualquier tipo de censura. De esta forma, queda claro que el Marco Civil no pretendía establecer una responsabilidad objetiva (teoría del riesgo) para los proveedores por hechos de terceros, debiendo determinarse cualquier responsabilidad a la luz de la responsabilidad subjetiva (teoría de la culpa)» (Teixeira, 2016, p. 110).

[63] «Artículo 57-F. El prestador de servicios de contenido y multimedios que hospede la difusión de propaganda electoral de candidato, partido o coalición será pasible de las sanciones previstas en esta Ley si, en el plazo determinado por la Justicia Electoral, contado a partir de la notificación de la decisión sobre la existencia de propaganda irregular, no tomar medidas para cesar tal difusión. Párrafo único. El prestador de servicio de contenido o multimedia sólo será responsabilizado por la difusión de propaganda si se comprueba que la publicación del material era de su conocimiento previo».

[64] Según el artículo 57-F, las plataformas que alojan propaganda electoral están sujetas a las sanciones previstas en la Ley Electoral si, en el plazo fijado

«impedir o disminuir» es extremadamente genérico, y conviene recordar que el principio de *lex certa es* una exigencia básica del derecho sancionador, de modo que se garantiza a las jurisdicciones una predeterminación normativa que deje claro cuáles son las conductas infractoras; y d) desde un punto de vista técnico, la norma es de difícil aplicación dada la enorme cantidad de casos límite y la inequívoca complejidad de los análisis semánticos en casos en los que intervienen, por ejemplo, «técnicas del silencio» (Grijelmo, 2012) y «verdades contrapuestas» (Macdonald, 2018), prácticas lingüísticas que dificultan enormemente la identificación de la desinformación.

En cualquier caso, entendemos que la imposición de un deber legal conduce a la aplicación, por analogía, del artículo 13, apartado 2, letras «a» y «c» del Código Penal[65], de modo que cualquier negligencia por parte de las plataformas ante resultados ilícitos puede ser apreciada como una «omisión relevante», a efectos de su penalización (con imposición de multa) en el ámbito de las representaciones.

En este sentido, el Reglamento establece una lista ilustrativa de medidas destinadas a señalar el cumplimiento del deber de

---

por la Corte Electoral, no adoptan medidas para poner fin a la difusión de contenidos ilegales. Sin embargo, la norma establece regímenes distintos para la actuación de los proveedores de contenidos (motores de búsquedas, redes sociales, plataformas de vídeos, *blogs*, etc.) y de los proveedores de información (portales de noticias en general). En el primer caso, la responsabilidad por contenidos de terceros sólo es posible cuando se acredita el conocimiento previo; en el segundo, el conocimiento previo es evidente, de modo que la sanción es independiente de la notificación previa para retirar el material irregular (Alvim, 2016, p. 326).

[65] «Artículo 13: El resultado, del que depende la existencia del delito, sólo puede atribuirse a la persona que lo ha causado. Se considera causa la acción u omisión sin la cual el resultado no se habría producido. [Párrafo 2: La omisión es penalmente relevante cuando el que omite debería y podría haber actuado para evitar el resultado. El deber de actuar incumbe a quienes: a) tienen una obligación legal de cuidado, protección o vigilancia; b) asumieron de otro modo la responsabilidad de evitar el resultado; c) con su comportamiento anterior, crearon el riesgo de que se produjera el resultado».

diligencia. Según los seis apartados del artículo 9-D[66], los objetivos regulatorios pueden alcanzarse, en principio, a través de: a) ajustes en la redacción y aplicación de condiciones de uso y políticas de contenidos; b) implantación de instrumentos de notificación eficaces y canales de denuncia abiertos a todos los usuarios y a instituciones o entidades públicas o privadas; c) planificación y ejecución de acciones correctivas y preventivas, incluyendo la mejora de los sistemas de recomendación de contenidos; d) transparencia de los resultados alcanzados por dicha mejora de los algoritmos de recomendación; e) elaboración de informes de impacto específicos, en año electoral, no sólo en términos de reducción de la desinformación contra las elecciones sino también de mitigación de los riesgos asociados; y f) mejora de las capacidades tecnológicas y operativas, priorizando herramientas y funcionalidades que contribuyan a los objetivos del artículo.

La adaptación de las condiciones de uso y las políticas comunitarias para reducir y evitar la desinformación antisistema es, según el punto I, un indicio de que la plataforma está actuando para cumplir con la obligación estipulada en el epígrafe. Cabe

[66] «Artículo 9-D. Es deber del proveedor de aplicaciones de Internet, que permite la difusión de contenidos político-electorales, adoptar y publicitar medidas para evitar o reducir la circulación de hechos notoriamente falsos o gravemente descontextualizados que puedan afectar a la integridad del proceso electoral, incluyendo: I – la elaboración y aplicación de condiciones de uso y políticas de contenidos compatibles con este objetivo; II – la implementación de instrumentos de notificación y canales de denuncia eficaces, accesibles a las personas usuarias y a las instituciones y entidades públicas y privadas; III – la planificación y ejecución de acciones correctivas y preventivas, incluyendo la mejora de sus sistemas de recomendación de contenidos; IV – la transparencia de los resultados alcanzados por las acciones mencionadas en el punto III del *caput* de este artículo; V – la preparación, en año electoral, de una evaluación del impacto de sus servicios en la integridad del proceso electoral, con el fin de aplicar medidas eficaces y proporcionadas para mitigar los riesgos identificados, incluso en lo que respecta a la violencia política de género, y la aplicación de las medidas previstas en el presente artículo; VI – mejorar sus capacidades tecnológicas y operativas, dando prioridad a las herramientas y funcionalidades que contribuyan a alcanzar el objetivo establecido en el encabezamiento de este artículo».

señalar, no obstante, que la mención a la «*observancia*» deja claro que la norma no se contenta con hipotéticas predicciones, por lo que el cumplimiento del deber de diligencia debe analizarse también desde la perspectiva de la *observancia,* reflejada en el análisis de la eficacia de los procesos de moderación. En otras palabras, la adopción de políticas adecuadas es una condición necesaria pero insuficiente: además, las consecuencias previstas en el plano de la moderación (etiquetado, marcado, reducción de la visibilidad, eliminación de contenidos, suspensión de cuentas, interrupción de la monetización o prohibición) deben aplicarse realmente, de modo que se promueva un amplio efecto disuasorio en el seno de las comunidades virtuales, en favor de la integridad de la información. Se trata aquí de evitar la práctica del *Ethics Washing* (blanqueamiento de imagen por motivos éticos) (Muñoz Vela, 2021, p. 61), por la que algunas empresas adoptan principios positivos como medida de relaciones públicas, sin intención de aplicarlos rígidamente.

En relación con el punto anterior, el punto II estipula que la apertura de canales para comentarios es un parámetro adicional del deber de cuidado, importante en la medida en que abre la perspectiva de depurar el debate público a instancias de los poderes públicos o de las organizaciones sociales, buscando así una forma de «gobernanza inclusiva» (Coeckelbergh, 2023, p. 157). La exigencia se justifica por el hecho de que la «curación» en un contexto estricto, interno y aislado, aunque sea proactiva, tiende a ser menos completa que la moderación en un contexto alimentado por un control multilateral, multitudinario y participativo. Sea como fuere, la norma exige «instrumentos eficaces», por lo que las denuncias deben tener efectos significativos. Esto significa que la apertura de canales de denuncia no cumple, por sí sola, el mandato: es necesario que las denuncias se tramiten con rapidez y eficacia, y que culminen con una resolución en plazo, para que el ciclo de moderación se realice adecuadamente.

La norma del punto III impone un cierto grado de profesionalidad, unido a la necesidad de que la lucha contra las falsas narrativas reciba, además de un esquema programático, un modelo de actuación anticipada, con fines de prevención. Las *medidas pre-*

ventivas pueden manifestarse, por ejemplo, en la realización de amplias campañas de concienciación o inmunización, así como en la contratación de agencias profesionales de verificación de hechos que ayuden a la moderación. También se plantea el uso de herramientas de IA para el análisis automatizado de grandes cantidades de información. Paralelamente, las propias adaptaciones de las normas de uso de la plataforma al contexto nacional pueden leerse como medidas cautelares. También se pueden realizar ajustes del producto, por ejemplo, adoptando precauciones que dificulten la propagación viral de contenidos desinformativos o nocivos. Por otro lado, pueden aplicarse *medidas correctoras*; por ejemplo, revisando los protocolos de registro, haciendo obligatoria la introducción de datos personales que creen obstáculos efectivos a la anonimización. E incluso realizar ajustes en las políticas y condiciones de uso de las comunidades, por ejemplo, excluyendo la monetización de los canales desinformativos, prohibiendo expresamente la desinformación contra las elecciones por parte de las plataformas que aún no lo hagan, o aumentando las sanciones por este tipo de prácticas para las plataformas que ya prohíben de algún modo las narrativas falsas contra las elecciones. Además, podrían ponerse en marcha investigaciones financiadas, promovidas o facilitadas por las propias plataformas, en colaboración con universidades o laboratorios de investigación que pongan a prueba la eficacia de la gobernanza interna, denunciando los puntos débiles y ayudando a revisar los protocolos para corregirlos.

En el punto IV, el imperativo de transparencia emerge como instrumento de supervisión y control, para que la adopción de medidas correctivas y preventivas no dependa exclusivamente de la autodeclaración de las *grandes tecnológicas*. En este contexto, las medidas adoptadas serán juzgadas a la luz de los respectivos resultados, según análisis cuantitativos y cualitativos que deberán poner de manifiesto una transformación positiva en la voluntad de cada plataforma de hacer sus entornos menos propicios y expuestos a la circulación de desinformación. En principio, y a título ilustrativo, desde un punto de vista *cuantitativo, se* podría informar, por ejemplo, del número de publicaciones prohibidas, cuentas suspendidas, contenidos etiquetados o marcados, canales

desmonetizados, así como de la cantidad de anuncios pagados irregulares detectados y bloqueados antes de su publicación; desde un *punto de vista cualitativo*, se podría informar, de la misma información que desde el punto de vista cuantitativo, del número de publicaciones prohibidas, cuentas suspendidas, contenidos etiquetados o marcados, canales desmonetizados, así como de la cantidad de anuncios pagados irregulares detectados y bloqueados antes de su publicación, la naturaleza, la frecuencia y el alcance de las campañas de sensibilización, añadiendo a este análisis los acuerdos celebrados con las instituciones públicas de control (el poder judicial, el Ministerio Fiscal, etc.), las asociaciones destinadas a difundir información adecuada sobre el funcionamiento de las elecciones, los ajustes de la normativa o de las prácticas de moderación, el desarrollo de nuevos productos como la inserción de etiquetas o *tags*, los ajustes para dificultar la propagación viral de contenidos ilícitos o para eliminar las cuentas anónimas, falsas o automatizadas, etc. En definitiva, la evaluación de impacto es un mecanismo de gobernanza capaz de mostrar los resultados de una actuación que se prevé positiva. En este contexto, su objetivo es hacer «demostrables» y «escalables» los beneficios obtenidos gracias a las medidas aplicadas.

La norma muestra que la Justicia Electoral espera mejoras no sólo en lo que se refiere específicamente a la mitigación de la desinformación y la neutralización de sus efectos, sino también en lo que se refiere a contenidos nocivos o abusos del lenguaje generalmente asociados a la difusión de *fake news* con ADN hostil o despectivo. Estas acciones, por tanto, pueden —y deben— abarcar otros fenómenos, como la violencia política de género, el discurso del odio[67], las prácticas de ciberacoso y las narrativas que incitan al extremismo y la radicalización.

[67] Óscar Sánchez, con fines explicativos, se ocupa de señalar las conexiones entre la desinformación y la incitación al odio: «Se trata de dos fenómenos distintos, pero entre los que existe un terreno común, pues, como se ha comprobado en los últimos años, la amplificación calculada del discurso de odio se ha convertido en una de las estrategias favoritas de las campañas de desinformación, para lo cual estas campañas se apoyan sobre las dinámicas sociales

A la luz del punto IV, se fomentan las innovaciones tecnológicas y operativas, que se contemplan previamente, como parámetros que demuestran el cumplimiento del deber de diligencia. Otros ejemplos hipotéticos de movimientos en esta dirección serían el desarrollo e implementación de nuevas tecnologías para detectar acciones coordinadas de desinformación o cuentas robot, medidas que favorezcan la trazabilidad de los mensajes y la preservación de mensajes temporales o autodestructivos, así como la prohibición de formas ocultas de publicidad y la creación de capas de refuerzo en la protección de datos personales. Además, el cumplimiento del deber de diligencia no tiene una forma definida, y puede adoptar la forma de una lista abierta de medidas no mencionadas en la resolución que, sin pretender ser definitiva, puede implicar, entre otras cosas: a) la promoción de iniciativas de alfabetización mediática, alfabetización informacional, alfabetización algorítmica e introducción a la IA; b) la adopción de medidas de gobernanza interna que fomenten la moderación de contenidos frente a mensajes desinformativos y nocivos; c) la implantación de procesos independientes de revisión y auditoría para reforzar *el cumplimiento* y garantizar la eficacia de los procedimientos internos, especialmente en lo que se refiere a *las ejecuciones* previas y en curso; d) la normalización de los procedimientos básicos de revisión de reclamaciones, notas y procedimientos relacionados, preferiblemente para reducir los tiempos de respuesta y actuación; y e) la introducción de elementos de transparencia, por ejemplo, estructurando las bibliotecas, habilitando APIs y proporcionando información y explicaciones sobre los flujos internos y la programación algorítmica (Muñoz Vela, 2022, pp. 95-96), con añadidos y adaptaciones.

Yendo más allá, el § 1 del artículo 9-D[68] prohíbe expresamente la promoción y priorización pagada de información desinfor-

existentes en las sociedades, agitando las divisiones y los conflictos internos» (Sánchez Muñoz, 2020, p. 33).

[68] «Apartado 1: Se prohíbe a los proveedores de aplicaciones que comercialicen cualquier forma de potenciación de contenidos, incluso en forma de priorización de resultados de búsqueda, que pongan este servicio a disposición

mativa contra la integridad del proceso electoral en los resultados obtenidos de las búsquedas de los usuarios, tanto en motores de búsqueda (Google, Bing u otros) como en otras plataformas de medios sociales (por ejemplo, búsquedas en plataformas de alojamiento de vídeos como YouTube, o redes sociales como Facebook, X o Instagram). Leído junto al § 2[69], queda claro que la obligación del § 1 tiene una naturaleza objetiva, estrictamente vinculada a los riesgos del negocio, que son innegablemente altos si se considera su posible impacto en el escenario electoral[70].

---

de la difusión de hechos notoriamente falsos o gravemente descontextualizados que puedan afectar a la integridad del proceso electoral».

[69] «Apartado 2. El proveedor de la aplicación que detecte un contenido ilícito de los referidos en el *caput* de este artículo o sea notificado de su circulación por los usuarios adoptará medidas inmediatas y efectivas para detener el *boosting*, la monetización y el acceso al contenido y promoverá la investigación interna del hecho y de los perfiles y cuentas implicados con el fin de evitar la ulterior circulación del contenido e inhibir conductas ilícitas, incluyendo la indisponibilidad del servicio de *boosting* o monetización».

[70] «Las redes sociales […] desempeñan un papel cada vez más influyente en el proceso electoral. Su capacidad para conectar a votantes, candidatos y organizaciones de noticias en tiempo real ha cambiado fundamentalmente la forma en que se llevan a cabo las campañas y cómo los votantes reciben la información. Twitter (ahora X), por ejemplo, se ha convertido en un foro clave del discurso político. Políticos, analistas y ciudadanos de a pie utilizan la plataforma para compartir sus opiniones, debatir temas y reaccionar a los acontecimientos en tiempo real. Ofrece a los candidatos una forma directa e inmediata de comunicarse con los votantes, pero también se ha utilizado para difundir información falsa e inflamar una retórica divisiva. Facebook, con su enorme base de usuarios y su sofisticada capacidad de segmentación publicitaria, ofrece una poderosa plataforma para difundir mensajes de campaña. Permite a los candidatos llegar a los votantes de forma muy personalizada, pero ha sido criticada por su falta de transparencia y su potencial de abuso. Instagram, especialmente popular entre los jóvenes, se utiliza a menudo para humanizar a los candidatos y atraer a los votantes de forma más visual y emocional. Sin embargo, también se ha utilizado para difundir desinformación a través de memes y otros contenidos compartibles. TikTok, aunque más reciente, ya ha demostrado un potencial significativo para influir en el discurso electoral. Sus vídeos breves y atractivos ofrecen una forma única de atraer a los votantes, especialmente a la generación más joven, pero también suscitan preocupación por la difusión de desinformación y contenidos manipuladores. En resumen, las redes sociales tienen el poder de

La norma, sin embargo, no se limita a prohibir la comercialización deliberada de servicios relacionados con el aumento artificial de la visibilidad de falsas narrativas contra el proceso electoral. Por el contrario, el incumplimiento se producirá siempre que las plataformas, incluso por descuido o negligencia, permitan la exposición privilegiada de contenidos falsos a cambio de una remuneración. De la combinación de ambas normas se deduce que las plataformas están obligadas a: a) no comercializar el impulso o la priorización de contenidos desinformativos en contra de la integridad electoral (§1); b) no permitir que se evite esta prohibición de impulso o priorización (§1); c) retirar inmediatamente los contenidos que hayan podido eludir los impedimentos, por propia iniciativa o provocación (§2, primera parte); y d) en caso de fallas, tomar medidas (promover cambios) para que no se repitan hechos similares (§2, *in fine*), perfeccionando los algoritmos de detección automatizada e intensificando el proceso de revisión humana, cuando proceda.

Además, el §2 amplía el abanico de prohibiciones al instituto de la monetización, creando una regla de desincentivación financiera de las actividades de desinformación. Así, corresponde a las plataformas no sólo bloquear el impulso, o la priorización, de contenidos de desinformación pagados, sino también impedir la remuneración de perfiles monetizados o canales dedicados a este tipo de actividad antisocial[71]. Según el texto, el fin de los pagos se

---

influir significativamente en el proceso electoral, tanto positiva como negativamente. Ofrecen nuevas vías para que los candidatos conecten con los votantes y para que éstos se comprometan con la política, pero también presentan nuevos riesgos y desafíos que requieren una atención y una regulación adecuadas. Es crucial que los votantes, los candidatos y los reguladores comprendan esta dinámica y trabajen para garantizar que los medios sociales se utilicen de forma que apoyen, y no socaven, el proceso democrático» (Lins, 2023, pp. 292-294).

[71] En particular, sin embargo, la resolución deja una laguna al no estipular el período durante el cual puede (o debe) suspenderse la monetización. La falta de un plazo mínimo, en este sentido, hace que las plataformas puedan dar por cumplida la norma establecida suspendiendo los pagos por un período irrisorio (24 horas, por ejemplo); la falta de un plazo máximo, en cambio, crea inseguridad jurídica y suele ser un elemento que induce a una fuerte judicialización.

impone *ex vi legis,* ya que las plataformas están obligadas a tomar tales medidas de forma proactiva, independientemente de que exista una denuncia social u oficial. La notificación por parte de los usuarios, en este sentido, sólo señala una inercia, que certifica —pero no «inaugura»— el contexto de violación de la norma impuesta. En términos directos, las plataformas asumen un deber de diligencia de impacto social, que les impide financiar a los participantes en la industria de la desinformación.

El incumplimiento conlleva la obligación de reparar el ecosistema informativo mediante la promoción gratuita de mensajes *que desmientan* las noticias falsas o las aclaren (por ejemplo, añadiendo información maliciosamente omitida u otros elementos contextuales)[72]. A título ilustrativo, la reparación —derivada de una suerte de derecho de réplica institucional anclado en el poder de policía— puede materializarse a través de notas aclaratorias, artículos de *fact-checking*, estudios especializados, documentos oficiales o informes públicos de cualquier tipo, a criterio de la autoridad judicial.

En este punto concreto, lamentablemente, la resolución no tuvo en cuenta que la desinformación, por su componente emocional, tiene una vocación natural de dispersión viral (Gillespie, 2020, p. 324), por lo que su alcance real, por regla general, supera con creces los límites del *boosting*, multiplicándose por innumerables reacciones de distribución social (*forwarding*) multiplataforma. En este sentido, se perdió la oportunidad de garantizar a los mensajes aclaratorios, de advertencia frente a la desinformación, una visibilidad más cercana a la que efectivamente obtienen los contenidos desinformativos, haciendo una analogía con el artículo 58, IV, b, de la Ley Electoral[73]. Cabe señalar que, en esta línea, consciente

[72] «Apartado 3. La Corte Electoral podrá ordenar al proveedor de la aplicación la difusión, vía *boosting* y de forma gratuita, de contenidos informativos que diluciden un hecho notoriamente falso o gravemente descontextualizado previamente *boosted* de forma irregular, en la misma forma y con el mismo alcance que el contrato».

[73] «Artículo 58: Una vez elegidos los candidatos en una convención, se garantiza el derecho de réplica al candidato, partido o coalición afectado, aunque sea indirectamente, por un concepto, imagen o declaración injuriosa, difamatoria, calumniosa o a sabiendas falsa difundida por cualquier medio de

de esta asimetría, el legislador ordinario prevé, en el artículo mencionado, que, en los casos de ofensa o divulgación de hechos no veraces en Internet, la respuesta estará a disposición de los usuarios durante no menos del doble del tiempo que estuvo disponible el mensaje ofensivo. En definitiva, la norma acaba dando a un fenómeno digital un tratamiento paralelo a las injusticias analógicas, cuando sería preferible adoptar la lógica del abuso de las libertades comunicativas en el escenario virtual. Sin embargo, no hay obstáculos para realizar esta corrección en el ámbito de la interpretación judicial.

4.2.4. Comportamientos perseguidos

La resolución traslada la lógica de la regulación asimétrica a la disciplina electoral, basada en la magnitud de los riesgos, algo muy típico, como hemos visto, en los documentos que regulan (o pretenden regular) la IA en todo el mundo. Los proveedores de aplicaciones, en este sentido, serán «responsables solidarios, civil y administrativamente, cuando no promuevan la indisponibilidad inmediata de contenidos y cuentas, durante el período electoral, en […] supuestos de riesgo» enumerados en los cinco apartados del artículo 9-E, que se refieren específicamente a: a) conductas, informaciones y actos antidemocráticos que tipifican los delitos previstos en los arts. 296, párrafo único (falsificación de sello o señal pública)[74], 359-L (abolición violenta del Estado Democrático

---

comunicación. […] IV – en la propaganda electoral en Internet: […] b) la réplica estará disponible para el acceso de los usuarios del servicio de Internet durante no menos del doble del tiempo que estuvo disponible el mensaje considerado ofensivo; […]».

[74] «[Falsificación de sello o signo público] Artículo 296 – Falsificarlos, fabricarlos o alterarlos: […] § 1 – Se castiga con las mismas penas: I – a quien haga uso del sello o signo falsificado; II – a quien utilice indebidamente el sello o signo verdadero en perjuicio ajeno o en beneficio propio o de terceros; III – a quien altere, falsifique o haga uso indebido de marcas, logotipos, siglas o cualesquiera otros símbolos utilizados por los órganos o entidades de la Administración Pública o que los identifiquen».

de Derecho)[75], 359-M (golpe de Estado)[76], 359-N (interrupción del proceso electoral)[77], 359-P (violencia política)[78] y 359-R (sabotaje)[79] del Código Penal brasileño; b) difusión o divulgación de hechos notoriamente falsos o gravemente descontextualizados que afecten a la integridad del proceso electoral, incluidos los procedimientos y técnicas de votación, el recuento y el cómputo de los votos; c) amenaza grave, directa e inmediata de violencia o incitación a la violencia contra la integridad física de los miembros y empleados de la Justicia Electoral y del Ministerio Público Electoral, o contra la infraestructura física del Poder Judicial, con el fin de restringir o impedir el ejercicio de los poderes constitucionales o la abolición violenta del Estado Democrático de Derecho; d) incitación al odio, incluida la promoción del racismo, la homofobia, las ideologías fascistas o de odio contra personas o grupos, en cualquier forma de discriminación; y e) difusión o intercambio de contenidos fabricados o manipulados, en parte o en su totalidad, por tecnologías digitales, incluida la IA, al margen del *descargo de responsabilidad* exigido.

En términos generales, la resolución impone a las plataformas el deber de excluir sumariamente aquellos contenidos que encajen en las hipótesis que describen un determinado conjunto de con-

---

[75] «[Abolición violenta del Estado Democrático de Derecho] Artículo 359-L. Intentar, con uso de violencia o amenaza grave, la abolición del Estado democrático de Derecho, impidiendo o restringiendo el ejercicio de los poderes constitucionales: […]».

[76] «[Golpe de Estado] Artículo 359-M. Intentar deponer, mediante violencia o amenaza grave, al gobierno legítimamente constituido: […]».

[77] «[Interrupción del proceso electoral] Artículo 359-N. Impedir o perturbar la elección o la verificación de sus resultados, violando indebidamente los mecanismos de seguridad del sistema de voto electrónico establecidos por la Justicia Electoral: […]».

[78] «[Violencia política] Artículo 359-P. Restringir, impedir u obstaculizar, mediante el uso de violencia física, sexual o psicológica, el ejercicio de los derechos políticos a cualquier persona por razón de su sexo, raza, color, etnia, religión u origen nacional: […]».

[79] «[Sabotaje] Artículo 359-R. Destruir o inutilizar medios de comunicación al público, establecimientos, instalaciones o servicios destinados a la defensa nacional, con el fin de abolir el Estado Democrático de Derecho: […]».

ductas delictivas vinculadas a la protección del régimen democrático, entre las que se incluyen, específicamente: a) la *falsificación de un sello o signo público*, que incluye el uso indebido de marcas, logotipos, siglas o cualesquiera símbolos identificativos de órganos o entidades de la Administración Pública (lo que, en el ámbito de la desinformación, se produce generalmente con la creación de cuentas falsas de órganos que forman parte de la Justicia Electoral); b) el *intento de abolir el Estado de Derecho,* con el uso de violencia o amenaza grave que impida o restrinja el funcionamiento de los poderes constitucionales; c) el *intento de deponer al gobierno electo*, mediante violencia o amenaza grave; d) *impedir o perturbar las elecciones o el recuento de sus resultados,* violando indebidamente los mecanismos de seguridad del sistema de voto electrónico; e) *restringir, impedir o perturbar el ejercicio de los derechos políticos, con el uso de* violencia física, sexual o psicológica, por razón de sexo, raza, color, etnia, religión u origen nacional; y f) *destruir o inutilizar medios de comunicación, instalaciones o servicios destinados a la defensa nacional, con el* fin de abolir el Estado Democrático de Derecho.

A pesar de la loable intención de salvaguardar el régimen de libertades públicas inherente al Estado constitucional, la norma presenta un aspecto muy preocupante, en la medida en que delega en las plataformas el deber de realizar un escrutinio extremadamente técnico que es innegablemente propio de las funciones del poder judicial[80], y además avanza en una dirección en principio contraria al régimen establecido por el artículo 19 del Marco Civil

[80] Cabe recordar que la legislación vigente establece que las *Big Techs*, por regla general, no son responsables por el contenido compartido por los usuarios en las redes sociales (artículo 19 del Marco Civil Brasileño de Internet). Según este esquema, la decisión de eliminar el contenido en caso de violación de los términos de uso o de las reglas de la comunidad es una mera opción de las plataformas (Lins, 2023, p. 299). Sin embargo, el Proyecto de Ley n.º 2.630/2020 (Proyecto de Ley de Fake News) pretende aprobar un régimen similar al establecido por el TSE en una reciente resolución. En caso de aprobación del proyecto, «las redes sociales serían responsabilizadas por contenidos que encuadren en algunos crímenes definidos en la legislación brasileña, como actos de terrorismo, instigación al suicidio o a la automutilación, crímenes contra el Estado

de Internet[81], que presupone un modo de responsabilidad restringido a la resistencia al cumplimiento de una orden judicial[82]. Además, tiende a ser altamente ineficaz, dado que la mayoría de los delitos mencionados incluyen la violencia, la amenaza grave o la vulneración de la seguridad del sistema de votación como elementos objetivos de los tipos delictivos, lo que los convierte en *delitos*

Democrático de Derecho, crímenes contra niños y adolescentes, violencia contra la mujer, entre otros» (Lins, 2023, *ibid*.).

[81] «La disposición parece ser una excepción a la regla del Marco Civil de la Internet (Ley 12.965/14) que, en su artículo 19, establece que, con el fin de garantizar la libertad de expresión y evitar la censura, el proveedor de aplicaciones de Internet sólo podrá ser considerado civilmente responsable de los daños y perjuicios derivados de contenidos generados por terceros si, tras una orden judicial específica, no adopta medidas para que los contenidos señalados como infractores no estén disponibles en el ámbito y límites técnicos de su servicio y en el plazo indicado, sin perjuicio de las disposiciones legales en contrario. A la vista de ello, existen serias dudas sobre si una Resolución del TSE podría ser el instrumento adecuado para ello, dada su naturaleza, y si no sería innovar e ir más allá de sus posibilidades» (De Teffé, 2024). Cabe señalar, sin embargo, que la constitucionalidad del artículo 19 está siendo cuestionada en el marco del Recurso Extraordinario n.º 1.037.396, ante el Supremo Tribunal Federal.

[82] Vale la pena destacar, sin embargo, el importante punto señalado por Ana Frazão: «[…] dado el estado actual de desarrollo de la inteligencia artificial, las experiencias pasadas y los riesgos que ya se han señalado, la iniciativa del TSE es esencial para garantizar la legitimidad del proceso electoral brasileño. Es por eso que incluso el argumento de una supuesta violación del artículo 19 del Marco Civil debe ser visto con cautela. Aparte del hecho de que se trata de una disposición legal cuya constitucionalidad está siendo discutida ante el STF y del hecho de que necesita ser interpretada de acuerdo con la legislación electoral, es esencial entender que […] el artículo 19 […] está claramente dirigido a contenidos de terceros, en relación a los cuales las plataformas asumen una posición de absoluta neutralidad. Contenidos sobre los cuales las plataformas tienen injerencia, muchas veces a través de la gestión y promoción, obviamente no pueden ser considerados como meros contenidos de terceros, razón por la cual no deben estar sujetos a las restringidas hipótesis de responsabilidad previstas en el artículo 19. En consecuencia, existen excelentes razones para justificar la compatibilidad de las Resoluciones del TSE con la Constitución y las leyes brasileñas. Sin embargo, la principal razón que las justifica es el simple hecho de que la democracia brasileña no puede esperar» (Frazão, 2024).

de acto, y no en *delitos de mera expresión,* que pueden realizarse mediante la simple enunciación de estos. En rigor, de acuerdo con la literalidad del texto, la norma obliga a las plataformas a retirar contenidos que se correspondan con los delitos enumerados, pero lo cierto es que cuatro de los seis delitos difícilmente podrían cometerse a través de manifestaciones en las redes sociales (las excepciones serían el uso indebido de símbolos y la perturbación del ejercicio del poder político mediante violencia psicológica, potencialmente observable en el discurso del odio y las campañas de acoso *online*). Desde el punto de vista del garante, en línea con el principio de seguridad jurídica, tal y como está redactado[83], el requisito no se extiende a las publicaciones que, aunque afecten a las instituciones democráticas, carezcan de cualquiera de los elementos necesarios para constituir uno de los delitos penales descritos.

Más bien, el deber de retirada debería dirigirse a la apología o incitación a esas prácticas, y no a las prácticas en sí mismas —y esto puede ser así si la interpretación se inclina por la expresión «casos de riesgo» contenida en el epígrafe, que equipararía la incitación pública a la comisión de los delitos en sí. De este modo, la norma daría cobertura efectiva a las publicaciones golpistas e insurreccionales, que sí han sido centro de preocupación desde que los «mensajes conspirativos» comenzaron a «pulverizar el consenso social», generando una absurda «polifonía de discursos» que «minaba la confianza en las instituciones públicas, pilar de las sociedades democráticas» (Carratalá; Iranzo-Cabrera; López-García, 2023, pp. 13-14). En esta coyuntura, sin embargo, una interpretación estricta y no expansiva, compatible con el dogma de las normas restrictivas del Derecho, indica que, técnicamente, las plataformas no están obligadas a eliminar los *posts* que no se

[83] «Artículo 9-E. Los proveedores de aplicaciones serán solidariamente responsables, civil y administrativamente, cuando no promuevan la indisponibilidad inmediata de contenidos y cuentas durante el período electoral, en los siguientes casos de riesgo: I – de conductas antidemocráticas, informaciones y actos que caractericen violaciones de los artículos 296, párrafo único; 359-L, 359-M, 359-N, 359-P y 359-R del Código Penal; […]».

ajusten a los tipos enumerados, lo que las exime de obligaciones respecto de *posts* especialmente relacionados con delitos cuya violencia o amenaza grave determine la comisión de un delito.

En cuanto al punto III, la ley acierta al incluir la perspectiva de la incitación a la violencia, ampliando el ámbito más allá de los casos de violencia propiamente dicha, que son nulos de pleno derecho. Sin embargo, la redacción reduce significativamente el alcance de la disposición al incluir la necesidad de un elemento subjetivo vinculado a la intención de «restringir o impedir el ejercicio de los poderes constitucionales o la abolición violenta del Estado democrático de Derecho». Idealmente, fomentar la violencia contra los órganos del Poder Judicial debería prohibirse, con independencia de la intención, a fin de incluir los llamamientos al vandalismo, que están más vinculados a un sentimiento general de revuelta que a una acción articulada dirigida a deponer a los gobiernos elegidos. Según el tenor literal de la norma, la incitación a atentar contra las sedes de las oficinas electorales difícilmente entraría en esta disposición, por lo que el mandato normativo, en muchas circunstancias, puede llegar a ser ineficaz.

La cruzada contra el odio y los prejuicios en el flujo de las conversaciones electorales, por otro lado, es urgente y beneficiosa, y responde al fin social de «protección de determinados colectivos, lucha contra la intolerancia y la discriminación», actualmente consagrado en el derecho internacional (Alcácer Guirao, 2023, pp. 31-32)[84]. Sin embargo, el punto IV del artículo 9 *sexies* suscita importantes preocupaciones en términos de seguridad jurídica y

[84] Por ejemplo, la Recomendación n.º 15 de la Comisión Europea contra el Racismo y la Intolerancia, publicada el 21 de marzo de 2016, promueve la lucha contra el discurso del odio. 15 de la Comisión Europea contra el Racismo y la Intolerancia, publicada el 21 de marzo de 2016, promueve la lucha contra el discurso del odio, que en este documento se caracteriza como «la promoción o instigación, en cualquiera de sus formas, del odio, la humillación o el menosprecio de una persona o grupo de personas, así como el acoso, el descrédito, la difusión de estereotipos negativos, la estigmatización o amenaza de una determinada persona o grupo de personas y la justificación de estas manifestaciones por motivos de raza, color, ascendencia, origen nacional o étnico, edad, discapacidad, lengua, religión o creencias, sexo, género, identidad de género, orien-

de la siempre complicada limitación de la libertad de expresión, precisamente en el punto en que inserta, junto a nociones jurídicamente determinables como racismo y homofobia, conceptos vagos que históricamente han sido objeto de importantes disputas semánticas, como nazismo y, especialmente, fascismo[85]. El exceso de abstracción, en particular, obliga a interpretaciones subjetivas que, a su vez, pueden dar lugar a casos de injusticia, tanto si la falta de igualdad ante la ley surge voluntariamente como si no. Además, la indeterminación semántica prácticamente imposibilitaría la acción automatizada, que sin duda es esencial para las acciones de defensa a gran escala.

Bien mirado, resulta extremadamente complicado definir los comportamientos nazi y fascista, más aún en una época en la que estos términos se utilizan a menudo de forma arbitraria y aleatoria, como armas retóricas destinadas a «construir significados mentales» (Castells, 2009, p. 535), con el objetivo de acumular poder y reivindicar el monopolio de la dignidad en una batalla de narrativas guiadas por «procesos de identificación y confrontación moral» (Elorza Saravia, 2023, p. 58), en la que el «significado emotivo de las nociones» hace que ciertas palabras sean «parte de la guerra» (Del Rey Morató, 2007, pp. 130-131).

En este sentido, cabe recordar que «el reino de los significados es muy complejo», dado que «los significados son variados, discrepantes y cambiantes» y pueden «ser imprecisos, abiertos a la interpretación». En las sociedades pluralistas, en particular, «no sólo son diferentes, sino que viven en tensión, competencia y conflicto», lo que impide aspirar a una «totalidad significativa», un significado unívoco para el conjunto de la sociedad. En última instancia, «todo significado es precario y problemático» (Martínez García, 2020, p. 30), especialmente en entornos fragmentados y polarizados, y puede ser cuestionado argumentalmente, dado que este tipo

---

tación sexual y otras características o condiciones personales» (Alcácer Guirao, 2023, p. 32). 32).

[85] Para más información sobre las disputas simbólicas en torno al concepto de fascismo, incluida la «corrupción chapucera de un término profundamente serio», véanse las observaciones de Mulhall (2021, pp. 29-31).

de categorizaciones estigmatizantes siempre serán objeto de discursos que movilizan valores, emociones, referencias y corrientes teóricas e ideológicas contrapuestas (Del Amo Castro, 2023, pp. 63-64). De este modo, cabe señalar que:

> «Los juegos de lenguaje de la ingeniería mágica permiten actuar sobre las nociones, ampliando o estrechando su significado. Se ensancha o restringe el campo de una noción de manera que englobe o no a ciertos seres, a ciertas cosas, a ciertas ideas, a ciertos comportamientos o a ciertas situaciones […] Hay nociones que son más plásticas que otras, que admiten ejercicios de estrechamiento o ensanchamiento del campo que abarcan, y se prestan a confusiones» (Del Rey Morató, 2007, pp. 129-130).

A pesar de estas consideraciones, lo cierto es que la norma aprobada, tal y como está redactada[86], tiene como objetivo central evitar la incitación al odio. En este contexto, los fenómenos asociados —«racismo», «homofobia», «ideologías nazis, fascistas o de odio», «prejuicios»— se mencionan desde una perspectiva claramente instrumental. Es decir, no necesitan ser vigilados, analizados y excluidos *per se,* sino sólo cuando se elaboran como *medios* (instrumentos) para la consecución de un *fin*: la práctica del discurso de odio, entendido en sentido amplio como cualquier forma de manifestación hostil o agresiva contra las minorías (Alcácer Guirao, 2023, p. 27), al hilo del «mayoritarismo de rebaño» que impregna las redes sociales (Fisher, 2023, p. 102). En cualquier caso, la vaguedad semántica recomienda desarrollar mejor el concepto.

---

[86] «Artículo 9-E. Los proveedores de aplicaciones serán solidariamente responsables, civil y administrativamente, cuando no promuevan la indisponibilidad inmediata de contenidos y cuentas, durante el período electoral, en los siguientes casos de riesgo: […] IV – de comportamiento o discurso de odio, incluyendo la promoción del racismo, homofobia, ideologías nazis, fascistas o de odio contra una persona o grupo por prejuicios de origen, raza, sexo, color, edad, religión y cualesquiera otras formas de discriminación; […]».

Se considera discurso de *odio* toda práctica lingüística dirigida a difundir afirmaciones u opiniones despectivas contra personas o grupos vulnerables o minoritarios (Alvim; Zilio; Carvalho, 2023, p. 182), precisamente por su condición vulnerable o minoritaria. En el ámbito de las cuestiones electorales, el discurso de odio implica un «deseo de exclusión», seguido de una acción dirigida a la perturbación personal, provocando, además de un sentimiento de no pertenencia, «la producción (mediante la incitación o la connivencia con la violencia) de daños a los adversarios», en un movimiento que interfiere en la normalidad y regularidad de la elección (Guaraty, 2023, p. 130). Dicho de otra manera:

> «El discurso de odio es una acción comunicativa que, basada en identidades sociales —muchas veces pertenecientes a minorías vulneradas o vulnerables—, intenta atacar la igualdad de las personas en dignidad y derechos, movilizando para ello diversas manifestaciones del lenguaje. Cuando hablamos de *discurso de odio* podemos pensar que se trata sólo de palabras, pero no es así. El discurso de odio trasciende a la comunicación verbal, sea ésta escrita o hablada. Hace uso de motivos visuales, símbolos, gráficos, imágenes y también de gestos o posiciones corporales que se vuelven parte de la composición, codificación y transmisión del mensaje, formas de utilizar nuestro idioma y nuestras ideas, que no son verbales. Todas las formas de comunicación podrían utilizarse como un medio para expresar un discurso de odio» (García Fajardo, 2022, pp. 243-244).

El discurso de odio, como hemos visto, puede llevarse a cabo de diversas formas, pero en cualquier caso la resolución opta por destacar algunas supuestamente más obvias o comunes. La noción de *prejuicio*, en primer lugar, se refiere a la idea de «creencias, opiniones o juicios de valor que carecen de justificación adecuada y suelen predisponer a las personas en contra de determinados grupos», lo que «provoca o perpetúa actitudes, situaciones o estructuras sociales injustas» (Stepanenko, 2022, p. 652). *El racismo*, paralelamente, implica «prácticas y discursos que distinguen, clasifican, discriminan y tratan de forma diferente a los seres huma-

nos, basándose en la idea de que existen entre ellos diferencias biológicas o corporales asociadas a su origen, capaces de justificar sistemas de dominación política y de explotación económica o científica» (Navarrete, 2022, p. 668)[87]. *La homofobia*[88], por su parte, toma forma cuando «el miedo o rechazo a los homosexuales se manifiesta a través de alguna forma de violencia» (Bustamante Tejada, 2022, p. 391)[89]. *La discriminación*, finalmente, emerge como el *leitmotiv* de todas las formas de abuso lingüístico previstas en el artículo. En términos conceptuales, puede entenderse como «un trato diferenciado que implica una distinción adversa derivada de valoraciones negativas realizadas sobre determinadas personas o grupos», generando situaciones que están «en

[87] «El racismo consiste en un comportamiento de odio, rechazo y desprecio hacia personas que supuestamente tienen características físicas diferentes (color de piel, tipo de pelo, etc.) y que se cree que forman una raza distinta y evidentemente inferior. El racismo pretende teorizar científicamente la existencia de estas razas consideradas claramente desiguales. Estas razas se conciben como especies animales jerarquizadas, cuyos rasgos inducen características morales y culturales particulares que controlan las acciones de cada individuo incluido dentro de cada raza» (Birnbaum, 2014, p. 254).

[88] «La desinformación ligada a las personas LGTBI ha experimentado un notable incremento en los últimos años alrededor del mundo, en países como USA, Polonia, Corea del Sur o Malasia, tanto vinculadas al debate público en período electoral (Rosinska, 2021), como resultado del tratamiento mediático […]. La circulación de estos mensajes desinformativos, que bien pueden ser considerados discurso de odio en tanto que persiguen generar daño a un grupo vulnerable, es mayor cuanto más resuenen con los valores morales dominantes en un contexto sociocultural determinado (Lelo y Caminhas, 2021) y, a menudo, ayuda a legitimar ciertas formas de gobierno y del ejercicio del control social mediante la creación de "pánicos sexuales" (Arbuet Osuna y Cáceres Soforza, 2019)» (Carratalá; Peris-Blanes, 2023, 240-241).

[89] A pesar de no señalarlo, está claro que los contenidos transfóbicos están cubiertos por la prohibición establecida en la resolución. Dicho esto, cabe señalar que «para las personas trans e intersexuales, la experiencia del odio y el maltrato es especialmente virulenta. Están incluso más expuestas a la discriminación extrema y a la violencia brutal que gays y lesbianas. Esto se debe, entre otras cosas, a la falta de espacios públicos donde puedan socializar y sentirse protegidos. En las piscinas, los vestuarios de los gimnasios o los aseos públicos, corren el riesgo de ser excluidos o agredidos» (Emcke, 2021, pp. 134-135).

contradicción directa con la idea de igualdad y justicia de las democracias constitucionales» (Luna Corvera, 2022, p. 238).

El reglamento añade que la supervisión proactiva de la actividad de las plataformas debe perseguir los contenidos que difundan discursos de odio envueltos en «ideologías nazis, fascistas o de odio contra una persona o grupo basadas en prejuicios de origen, raza[90], sexo, color, edad, religión y cualquier otra forma de discriminación», como la orientación sexual, la identidad de género, el origen étnico, la discapacidad, etc., para incluir discursos fóbicos de todo tipo. Sin embargo, en lo que se refiere a las cosmovisiones, no se trata, como ya se ha aclarado, de definir lo que constituye —o no— un discurso que lleva trazas de ideologías destructivas, sino de saber si, en torno a cualquier idea nazi-fascista que se haga pública, hay un discurso de odio incrustado. La reducción de la complejidad, a partir de esta premisa, es evidentemente brutal, lo que facilita enormemente la tarea del operador jurídico.

Aun así, no es necesario explicar que las intersecciones históricas entre *fascismo* y discurso de odio son menos evidentes que el vínculo intrínseco entre *nazismo*, racismo y persecución discriminatoria (Borja, 1997, p. 704), sobre todo porque «el énfasis que puso en el racismo y el antisemitismo» es precisamente la «prin-

[90] «Según el sociólogo [Oracy Nogueira], el prejuicio racial es "una disposición (o actitud) desfavorable, culturalmente condicionada, hacia miembros de una población que son percibidos como estigmatizados, ya sea por su apariencia o por toda o parte de la ascendencia étnica que se les atribuye o reconoce". A partir de este concepto, designa dos tipos de prejuicio: el prejuicio de marca y el prejuicio de origen, utilizando los ejemplos de las relaciones sociales en Estados Unidos y Brasil para calificar uno y otro. Si existe un prejuicio basado en los rasgos físicos, los gestos o el acento de una persona, se trata de un prejuicio de marca, que varía subjetivamente en función de las características de la persona que observa y de la persona identificada. Cuando la suposición de que una persona desciende de un determinado grupo étnico-racial basta para convertirla en víctima de un prejuicio, se trata de un prejuicio de origen. En las relaciones raciales definidas por el prejuicio de marca, el criterio para definir los grupos discriminador y discriminado es la apariencia racial y en las relaciones raciales definidas por el prejuicio de origen, independientemente de la apariencia o proporción de ascendencia del grupo discriminador o discriminado, si hay mestizaje habrá prejuicio» (Braga, 2022).

cipal diferencia» entre ambas visiones del mundo (Costa, 2020, p. 293). Sin embargo, en algunas experiencias, el fascismo, en sus conexiones con el nacionalismo exacerbado (Saccomani, 2009, p. 466), emana impulsos xenófobos (Pérez-Curiel; Rivas-de-Roca; García-Gordillo, 2023, p. 29) y difunde la intolerancia religiosa. En su versión posmoderna, por cierto, se reconoce que el fascismo, en su tendencia a señalar enemigos, suele difundir cadenas de ataques digitales no sólo contra extranjeros e islamistas (o contra adeptos de religiones de origen africano, en el caso de Brasil), sino también contra homosexuales (Carratalá; Peris-Blanes, 2023, p. 24), defensores del islam (Carratalá; Peris-Blanes, 2023, p. 24). 241), defensores de la ideología de género (Montagut; Willem; Carrillo, 2023, p. 256) y activistas feministas (Murgia, 2019, p. 49), que reflejan agendas xenófobas e islamófobas (Emcke, 2021, p. 55; Mulhall, 2022, p. 25), agendas transfóbicas y misóginas típicas de los movimientos *alt-right*, particularmente en el campo de la extrema derecha radical (Massanari, 2020, p. 179)[91].

4.2.5. Efectos sobre la libertad de expresión

Dicho esto, no se puede descartar el efecto limitador que esta norma puede tener sobre el clima en el que se desarrolla la libertad de expresión, dado que, desde el punto de vista de las grandes empresas, las leyes con este alcance fomentan la eliminación de todo tipo de contenidos controvertidos, incluidos los legítimos, por miedo a ser sancionados (Lins, 2023, p. 300), lo que evidentemente puede repercutir tanto en el ámbito de la libertad de expresión

[91] «La jerarquía, que puede estar relacionada con la raza, la etnia, la condición sexual, la posición política, está siempre presente en las ideologías de extrema derecha, y la radicalización, cuando es llevada al extremo, produce un ambiente de persecución contra aquellos que, dentro de sus creencias, son seres humanos inferiores y que no deberían gozar de los mismos derechos y protecciones de las libertades individuales. Por lo general, la retórica de la extrema derecha hace explícita la tendencia a deshumanizar a los grupos que considera inferiores» (Prado, 2023, p. 80).

de opiniones e ideas como, hipotéticamente, en la igualdad de condiciones entre candidatos y partidos políticos, ya que la moderación de contenidos, mal realizada, puede generar desequilibrios. Sin embargo, como se preguntan Henry Kissinger y sus colaboradores: dentro de la sociedad constitucional, ¿existe un «derecho a leer [producir o difundir], o incluso un interés legítimo en leer, información "falsa" generada por la IA?» (Kissinger; Schmidt; Huttenlocher, 2023, p. 106).

Es más, cabe preguntarse si la exigencia de eliminar los contenidos generados artificialmente que no cuenten con *disclaimers* sobre su origen sintético es factible desde un punto de vista técnico. Ello se debe a que el enorme flujo de publicaciones diarias hace imprescindible el análisis automatizado, de modo que el cumplimiento de la norma requeriría una combinación de herramientas de análisis semántico (operando, como hemos visto, sobre términos excesivamente vagos) y de *software* de tratamiento de imágenes o de computación visual, posiblemente combinadas con la intervención posterior de revisores humanos, movilizados para minimizar los errores potencialmente cometidos por las máquinas, por ejemplo, cuando los términos relacionados con falsas narrativas son objeto de artículos periodísticos genuinamente informativos (que, por ejemplo, se hacen eco de decisiones condenatorias), o de manifestaciones que los reproducen con el único fin de negarlos o criticarlos públicamente.

Por último, es conveniente destacar los esfuerzos de armonización en esta materia. La justicia electoral brasileña cuenta con más de 2.600 magistrados que actúan en las zonas electorales. Considerando la necesidad de una actuación uniforme, a fin de tratar casos equivalentes por igual, se ha establecido que las decisiones colegiadas del TSE serán vinculantes a las actuaciones de primera instancia, tanto en los procedimientos administrativos basados en el poder de policía como en la resolución de demandas (derecho de réplica o publicidad irregular)[92].

[92] Resolución TSE n. 23.610/2019 – «Artículo 9-F. En caso de que la propaganda electoral en Internet transmita hechos notoriamente falsos o gravemente descontextualizados sobre el sistema de voto electrónico, el proceso electoral o

En este contexto, la resolución estipula que el trato equivalente abarca no sólo los casos de contenido idéntico, sino también aquellos en los que existe un estado de «similitud sustancial» entre el contenido original y la réplica aproximada[93]. De hecho, la norma aborda el hecho de que el concepto de «identidad» es difícil de determinar en la práctica. Después de todo, en un número casi infinito de publicaciones sobre el mismo tema, ¿qué puede considerarse, en un análisis riguroso, objetivamente idéntico? Obviamente, en una interpretación autónoma, la expresión «idéntico» difiere claramente de lo que es «análogo», «correlativo» o simplemente «similar». Sin embargo, teniendo en cuenta la complejidad y el alto grado de variación entre los casos, el Tribunal Supremo optó por equiparar «idéntico» con «sustancialmente similar» con el fin de hacer la norma más operativa.

Por tanto, es posible deducir que la regla debe aplicarse no sólo a contenidos idénticos al 100% (lo que implica la inexistencia absoluta de diferencias entre la pieza original y la pieza derivada), sino también —y como mínimo— a los casos de: a) reproducción parcial, especialmente mediante la exhibición de recortes aislados o fragmentos de «alto impacto»; b) repetición en un soporte trasplantado (difusión de audio aislado de un vídeo considerado desinformativo); c) edición simple (aceleración del tiempo de palabra,

---

la Justicia Electoral, los jueces mencionados en el artículo 8.º de la presente Resolución se sujetarán, en ejercicio del poder de policía y en las representaciones, a las decisiones del plenario del Tribunal Superior Electoral sobre la misma materia, en las que se haya ordenado la remoción o el mantenimiento de contenido idéntico» (disponible en: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019], consultado el: 4-5-2024).

[93] Resolución TSE n. 23.610/2019 – Artículo 9-F. «Párrafo 1.º Lo dispuesto en el cuerpo principal de este artículo se aplicará a los casos en que, a pesar de la edición, reestructuración, alteración de palabras u otros artificios, métodos o técnicas para eludir los sistemas automáticos de detección de contenido duplicado o para dificultar la verificación humana, exista similitud sustancial entre el contenido removido por orden del Tribunal Superior Electoral y el difundido en la publicidad regional o municipal» (disponible en: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019], consultado el: 4-5-2024).

inserción de subtítulos, etc.); d) edición compleja (aprovechamiento de momentos clave para crear nuevos contenidos con la misma base simbólica, por ejemplo.); d) montaje complejo (aprovechamiento de momentos clave para crear nuevos contenidos con la misma base simbólica, por ejemplo, con la adición de elementos sonoros para hacerlo más dramático, divertido, etc.) y; e) superposición (por ejemplo, en los contenidos *reactivos,* donde el contenido glosado se reproduce en *segundo plano,* mientras es analizado o reforzado por las *apreciaciones* realizadas por un comentarista).

La cuestión, sin embargo, tiene otras capas, y las hipótesis de derivación (*spin-offs*) son particularmente complejas, por ejemplo, con el uso de transcripciones parciales, como *hashtags,* tarjetas, pies de foto o memes que reproducen, en texto, una idea-fuerza presente en una pieza considerada desinformativa. A modo de ejemplo: una vez que el Alto Tribunal ha considerado desinformativa la propaganda oficial del partido X que dice que el candidato Y es un asesino, ¿pueden considerarse sustancialmente similares las futuras publicaciones con la misma acusación, en todos los casos? Además, la complejidad aumenta cuando recordamos la importancia del contexto, sobre todo porque en muchos casos la reproducción puede estar vinculada a una actividad periodística, humorística, etc., o incluso desde una perspectiva crítica a la publicación original (reproducción con fines de censura, contradicción, sensibilización, etc.). Por tanto, la comparación de vídeos u otras piezas debe tener en cuenta no sólo su identidad o similitud temática, sino también su intencionalidad y el marco contextual.

En suma, el edicto reglamentario establece la creación de un repositorio de decisiones colegiadas[94] creado para facilitar el tra-

[94] Resolución TSE n. 23.610/2019 – «Artículo 9-G. Las decisiones del Tribunal Superior Electoral que ordenen la retirada de contenido que transmita hechos notoriamente falsos o gravemente descontextualizados que afecten a la integridad del proceso electoral se incluirán en un repositorio puesto a disposición para consulta pública. § Párrafo 1.º – El repositorio contendrá el número de la causa y el texto completo de la decisión, de la cual se destacará la dirección electrónica donde está alojado el contenido a ser removido y una descripción de sus elementos esenciales para inclusión en un campo proporcionado por la Secretaría Judicial. […]» (disponible en: [https://www.tse.jus.br/legislacao/com-

bajo de las autoridades zonales, además de definir que las órdenes de retirada de contenido basadas en la existencia de una manifestación colectiva del Tribunal Superior podrán indicar un plazo para su cumplimiento inferior a 24 horas, dependiendo de la gravedad del contenido y del contexto de las elecciones en curso (artículo 9-F, §§ 2 y 3). Sin embargo, con el ánimo de evitar abusos y garantizar así la posición preferente que ocupa la libertad de expresión en el fuero constitucional, crea un mecanismo destinado a promover la autocontención del Poder Judicial, atenuando los excesos en el ejercicio del poder de policía (artículo 9-F, § 4[95]). Esta norma abre la posibilidad de que las decisiones ilegales o desproporcionadas (abusivas) sean impugnadas a través de reclamaciones, lo que puede llevar, acumulativa o alternativamente, a la anulación o revocación de la orden y a la comunicación del hecho al respectivo Departamento de Asuntos Internos, para el inicio de una investigación o de un procedimiento administrativo disciplinario.

Para que las decisiones de retirada de contenidos desinformativos se adopten con objetividad y seguridad, las autoridades judiciales deben actuar con moderación y abstenerse de ordenar la retirada de mensajes sin tener la seguridad sobre su veracidad. También, se recomienda que actúen de acuerdo con criterios técnicos, observando lo que la doctrina ha denominado las «condiciones de legitimación para la restricción judicial de contenidos», que implican, por un lado, requisitos *negativos* (que no pueden estar presentes) y *requisitos positivos* (cuya incidencia es necesaria para la legitimación). Según Frederico Alvim, Rodrigo Zilio y Volgane Carvalho, los requisitos negativos son: a) la ausencia de

---

pilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019], consultado el: 4-5-2024).

[95] Resolución TSE n. 23.610/2019 – Artículo 9-F – «Párrafo 4.º El ejercicio del poder de policía que contradiga o exceda lo dispuesto en el parágrafo 1.º de este artículo permitirá el uso de la reclamación administrativa electoral, con sujeción a lo dispuesto en los arts. 29 y 30 de la Res. TSE 23.608/2019» (disponible en: [https://www.tse.jus.br/legislacao/compilada/res/2019/resolucao-no-23-610-de-18-de-dezembro-de-2019], consultado el: 4-5-2024).

duda objetiva o estado de indeterminación; b) la ausencia de afirmaciones que no estén sujetas al régimen de la verdad fáctica (como las manifestaciones artísticas y religiosas, por ejemplo); c) la ausencia de afirmaciones que no puedan ser verificadas; y d) la ausencia de transparencia en cuanto al carácter no veraz de las afirmaciones. Los requisitos *positivos*, por su parte, se refieren a: a) la conciencia de la falsedad (fáctica o contextual); b) la lesividad expresiva; c) el alcance relevante; y d) la intencionalidad lesiva (que eventualmente puede superar las dudas sobre la presencia del primer requisito). Además, los autores citados advierten que:

> «[…] en un escenario coherente con la garantía de un debate público libre y plural, no hay razones para cercenar la *crítica ácida basada en premisas correctas,* ni para obstaculizar las *afirmaciones erróneas no rotundas* (con alcance limitado). Del mismo modo, debe valorarse la intención de engañar, ya que en muchas circunstancias la mera *burla* o los *errores involuntarios* no merecen atención judicial. […] el recorte de la libertad de expresión no se justifica ante *acusaciones inocuas* o *elucubraciones banales, basadas en el señalamiento* de hechos o circunstancias intrascendentes o inocuas. Los juicios morales participan naturalmente de la confrontación política, y las ofensas liliputienses, por ser inocuas y corrientes, no deben ser tratadas como anomalías» (Alvim; Zilio; Carvalho, 2023, pp. 323-324).

Además, se estipula que, en el plazo fijado, las plataformas digitales no sólo deben cumplir las órdenes de retirada, sino también, cuando se determine específicamente, contribuir a alimentar el repositorio, indicando: a) el archivo de texto, imagen, audio o vídeo objeto de la orden de retirada; b) las capturas de pantalla que contengan todos los comentarios disponibles en el sitio de alojamiento del contenido; c) los metadatos relativos al acceso, como IP, puerto, fecha y hora de publicación; y d) los metadatos relativos a la participación de la publicación en el momento de su retirada (artículo 9.º-G, § 2.º).

La orden impone cargas adicionales a las plataformas digitales que, por regla general, no sólo deben cumplir las órdenes, sino

también informar de su efectivo cumplimiento. Igualmente, faculta a las autoridades judiciales para que, mediante orden expresa, obliguen a los proveedores de aplicaciones de Internet a alimentar el repositorio creado, incorporando una cantidad considerable de información. El problema es que, teniendo en cuenta la magnitud de las elecciones en determinados países (recuérdese que, por ejemplo, el contexto brasileño implica a más de 2.600 autoridades en acción), las cargas parecen desproporcionadas y tal vez contraproducentes, ya que la atención de las plataformas se divide: en lugar de centrarse principalmente en la rápida eliminación de los *posts* con contenido nocivo, se dedicarán al mismo tiempo al cumplimiento (posiblemente lento) de tareas burocrático-administrativas.

Por si fuera poco, fijar plazos evidentemente cortos para alimentar el repositorio podría hacerlo inaplicable, poniendo a las plataformas en un estado permanente de rebeldía involuntaria. En cualquier caso, cabe esperar que los datos relevantes, como la tabla de comentarios (importante para demostrar la nocividad y repercusión social de los mensajes retirados) y la información sobre *engagement* (útil para certificar el alcance y grado de malestar generado por los contenidos retirados) sean efectivamente utilizados por los organismos de control y las entidades productoras de conocimiento, permitiendo comprender y combatir mejor el fenómeno de la desinformación. Sin olvidar que los gigantes tecnológicos obtienen ingresos más que suficientes para reforzar y adaptar sus equipos, por lo que la tarea, aunque compleja y difícil de cumplir, no es en absoluto desproporcionada.

El artículo 9-H va más allá y reproduce una pauta normal para el ordenamiento jurídico brasileño, estableciendo la lógica del *non bis in idem,* explicando que la efectiva remoción de contenido no exonera las responsabilidades de los usuarios involucrados, que posteriormente pueden ser multados en el ámbito de las representaciones electorales. En este sentido, la supresión de contenidos no impediría el establecimiento acumulativo de sanciones pecuniarias, siempre y cuando se observe la garantía de un proceso contradictorio y la plena defensa en un juicio posterior.

En un punto más avanzado, el TSE, muy atento a la orientación al beneficio que guía las decisiones y el comportamiento de las plataformas —pese a la imagen que proyectan a través de incisivas actividades de relaciones públicas, siempre marcadas por una «retórica laudatoria» (Morozov, 2018, p. 48)—, condiciona el ánimo de lucro a un marco adicional de cautelas, especialmente válido para las empresas que se lucran con la elaboración de perfiles de usuarios y la consiguiente oferta de potenciación de contenidos. En este segmento, incorpora recomendaciones que también se incluyen en un informe elaborado por la Comisión de Venecia sobre los principios fundamentales para preservar la integridad ante las nuevas tecnologías digitales (Comisión de Venecia, 2020, p. 8).

4.2.6. Transparencia y protección de datos

El artículo 27-A[96] establece que los proveedores que exploten el mercado de amplificación artificial de mensajes en medios digitales, incluso mediante la priorización de resultados de búsqueda, deberán mantener un repositorio completo de anuncios, dotado de un mecanismo de consulta eficaz que permita una fácil

[96] «Artículo 27-A. El proveedor de aplicaciones que preste el servicio de potenciación de contenidos político-electorales, incluso en la forma de priorización de resultados de búsqueda, deberá: I – mantener un repositorio de estos anuncios para monitorear, en tiempo real, el contenido, los montos, los responsables del pago y las características de los grupos poblacionales que conforman la audiencia (*profiling*) de la publicidad contratada; II – proporcionar una herramienta de consulta, accesible y de fácil uso, que permita realizar búsquedas avanzadas sobre los datos del repositorio, que contenga, como mínimo: (a) búsquedas de anuncios basadas en palabras clave, términos de interés y nombres de anunciantes; (b) acceso a información precisa sobre los importes gastados, el período de impulso, el número de personas alcanzadas y los criterios de segmentación definidos por el anunciante en el momento de servir el anuncio; (c) recopilaciones sistemáticas, mediante una interfaz dedicada (interfaz de programación de aplicaciones – API), de los datos de los anuncios, incluyendo su contenido, gasto, alcance, audiencia alcanzada y los responsables del pago».

supervisión social del panorama de la financiación de la circulación de contenidos políticos[97]. Como resultado, la sociedad, los órganos académicos, la prensa, los competidores y los órganos de control podrán supervisar, en tiempo real, los actores, valores y datos utilizados en la distribución de mensajes perfilados, a través de búsquedas realizadas por simple *scraping* (búsqueda de anuncios por palabras clave, términos de interés o nombres de anunciantes), o de forma automatizada y profesional, a través de APIs (*interfaces de programación de aplicaciones*).

De ello se desprende que los importes invertidos, el período contratado, el número de usuarios alcanzados y los criterios psicométricos o sociodemográficos utilizados para definir el público objetivo deberán ser públicas, lo que conferirá a las competiciones electorales brasileñas una transparencia probablemente sin precedentes en términos mundiales.

Estas medidas, que se aplican no sólo durante el período oficial de campaña, sino de forma permanente (artículo 27-A, II), deben aplicarse en un plazo máximo de dos meses a partir de la entrada en vigor de la resolución (en el caso de plataformas ya establecidas en Brasil), o desde el inicio de la prestación de servicios, en el caso de futuros proveedores que comiencen a ofrecer sus servicios en el escenario brasileño. La adecuación de conducta, en particular, es vista por la norma como condición obligatoria para la acreditación ante la Justicia Electoral y, consecuentemente, como requisito indispensable para la obtención de ingresos por actividades de esta naturaleza (artículo 27-A, § 4, de la Res. TSE n.º 23.610/2019).

Además, los cambios introducidos por la Resolución 23.723/2024 en la Resolución de Publicidad (Res. TSE n. 23.610/2019) también representan un hito significativo en los esfuerzos relacionados con

[97] «Párrafo 1.º A los efectos de este artículo, se caracterizan como contenidos político-electorales, independientemente de la clasificación que haga la plataforma, aquellos que versen sobre elecciones, partidos políticos, federaciones y coaliciones, cargos electivos, personas que ocupan cargos electivos, candidatos, propuestas de gobierno, proyectos de ley, ejercicio del derecho de sufragio y demás derechos políticos o materias relacionadas con el proceso electoral».

la protección de datos personales, con el fin de garantizar que el uso de tecnologías emergentes se produzca de manera responsable y transparente en el entorno de las campañas electorales. Esta acción es especialmente relevante dado que ni la legislación brasileña de protección de datos ni la legislación electoral han tratado expresamente el tratamiento de datos en el contexto de las elecciones (Souza, 2022, p. 78[98]). Aunque hubiera sido preferible que la materia fuera regulada en detalle por el Poder Legislativo, entre otras cosas por el carácter más democrático de su tarea normativa, el modelo institucional brasileño ha dotado al Tribunal Superior Electoral de la potestad reglamentaria para dictar resoluciones destinadas a salvaguardar el buen desarrollo de las elecciones (Souza, 2022, pp. 81-82).

Para ello, el §4 del artículo 10 establece que el tratamiento de los datos debe ajustarse a la finalidad para la que fueron recogidos, y que deben observarse los principios y reglas de la Ley n.º 13.709/2018 —Ley General de Protección de Datos (LGPD)—[99].

---

[98] El autor menciona algunas disposiciones escasas que abordan la cuestión de forma tangencial, como los artículos 57-E y 57-G de la Ley Electoral (Ley n.º 9.504/97), que tratan de la prohibición del uso y donación de registros electrónicos, y del derecho del interesado a exigir que se interrumpan las comunicaciones políticas (Souza, 2022, pp. 78-79).

[99] Cuando las herramientas de IA toman las riendas de las campañas, traen consigo un conjunto de postulados éticos y normativos pertinentes a la conducta guiada de la propia tecnología, en conexión con el marco de tratamiento de los derechos y deberes de la sociedad digital. Esto es lo que ocurre, por ejemplo, con las normas establecidas en las leyes de protección de datos y regulación de las aplicaciones de Internet. Además, y como es obvio, están sometidas a la disciplina estructurante del Derecho Electoral, que incluye el ineludible sometimiento a una cartera de principios generales y específicos. Básicamente, estas premisas parten de la idea básica de que las innovaciones tecnológicas, en principio, no requieren un tratamiento completo y fundacional, siempre que los estatutos de protección puedan ser ajustados, por vía interpretativa, para mantener la coherencia del sistema y el cumplimiento de su finalidad social. Como resume Lucrecio Rebollo, en medio de este escenario «la protección jurídica y los medios de garantía ya existen, lo que varía son las formas de vulneración, por ello no es necesario crear un ordenamiento jurídico *ex novo*, es necesario ir adecuando el existente a las nuevas necesidades, y ello es tarea tanto del le-

En consonancia con esta ley, la resolución (artículo 10, §5) estipula que los agentes electorales deben proporcionar información clara sobre el tratamiento de los datos personales, así como crear canales para que los votantes ejerzan sus derechos individuales, como solicitar la eliminación o la baja en las bases de datos (artículo 10, §5). Esta medida refuerza el control de los individuos sobre su información en el contexto de las competiciones políticas (autodeterminación informativa). En el mismo sentido, el §9 del artículo 28 establece que la propaganda electoral que implique el tratamiento de datos personales sensibles deberá cumplir con las condiciones previstas en la LGPD[100].

El reglamento también establece que, en las elecciones municipales, en ciudades con menos de 200.000 electores, los partidos políticos y entidades afines (coaliciones y federaciones), así como los candidatos, serán considerados pequeños agentes de tratamiento, teniendo en cuenta lo dispuesto en la Resolución 2/2022 de la

---

gislador, de la jurisprudencia y también de la doctrina» (Rebollo Delgado, 2023, p. 52).

[100] Ley n.º 13.709/2018: «Artículo 11. Tratamiento de datos personales sensibles. El tratamiento de datos personales sensibles sólo podrá tener lugar en los siguientes casos: I – cuando el titular de los datos o su representante legal consienta, de forma específica y separada, para fines específicos; II – sin proporcionar el consentimiento del titular de los datos, en los casos en que sea indispensable para: a) el cumplimiento de una obligación legal o reglamentaria por parte del responsable del tratamiento; b) el tratamiento compartido de datos necesarios para la ejecución, por parte de la administración pública, de políticas públicas previstas en leyes o reglamentos; c) la realización de estudios por parte de un organismo de investigación, garantizando, siempre que sea posible, la anonimización de los datos personales sensibles; d) el ejercicio regular de derechos, incluso en un contrato y en procedimientos judiciales, administrativos y de arbitraje, estos últimos en lo/.307, de 23 de septiembre de 1996 (Ley de Arbitraje); e) protección de la vida o de la seguridad física del interesado o de un tercero; f) protección de la salud, exclusivamente en procedimientos realizados por profesionales de la salud, servicios de salud o autoridades sanitarias; o g) garantía de la prevención del fraude y de la seguridad del interesado, en los procesos de identificación y autenticación de registro en sistemas electrónicos, salvaguardando los derechos mencionados en el artículo 9 de esta Ley y salvo en el caso de que prevalezcan los derechos y libertades fundamentales del interesado que requieran la protección de datos personales».

Agencia Nacional de Protección de Datos (ANPD). Así, en los términos del artículo 10, §6-B, a) están exentos de designar al responsable del tratamiento de los datos personales; y b) tienen derecho a establecer una política simplificada de seguridad de la información, «que deberá incluir los requisitos esenciales y necesarios para el tratamiento de los datos personales, con el objetivo de protegerlos del acceso no autorizado y de la destrucción, pérdida, alteración, comunicación o cualquier forma de tratamiento inadecuado o ilícito, accidental o ilícito».

El nuevo artículo 33-B asigna a los proveedores de aplicaciones y a los actores electorales algunas obligaciones específicas cuando tratan datos, a saber: a) garantizar el fácil acceso a la información del tratamiento, especialmente a los datos utilizados para la elaboración de perfiles de usuarios, como medida dirigida a la publicidad micro-segementada[101]; b) garantizar a los afectados los derechos previstos en los arts. 17 a 20 de la LGPD (acceso a los datos, anonimización, revocación del consentimiento, etc.); c) adoptar las medidas necesarias para proteger contra la discriminación ilícita y el abuso; d) abstenerse de utilizar los datos para fines distintos de los explícitamente consentidos; c) adoptar las medidas necesarias para proteger a los usuarios contra la discriminación ilícita y abusiva; d) abstenerse de utilizar los datos para fines distintos de los explicados y consentidos por los respectivos interesados, de conformidad con los principios de finalidad, necesidad y

[101] De acuerdo con el glosario incluido en el artículo 37 de la misma resolución, se entiende por perfilación el «tratamiento de múltiples tipos de datos de personas físicas identificadas o identificables, realizado generalmente de manera automatizada, con la finalidad de conformar perfiles con base en patrones de conducta, gustos, hábitos y preferencias y clasificar dichos perfiles en grupos y sectores, utilizándolos para análisis o predicciones de movimientos y tendencias de interés político-electoral» (punto XXXII). El micro-direccionamiento, por su parte, «es la estrategia de segmentación de la propaganda electoral o comunicación de campaña que consiste en seleccionar a personas, grupos o sectores, clasificados mediante la elaboración de perfiles, como público objetivo o audiencia de los mensajes, acciones y contenidos político-electorales elaborados con base en los intereses perfilados, con la finalidad de incrementar la influencia en su comportamiento» (punto XXXIII).

adecuación; e) aplicar ajustes de seguridad para proteger contra el acceso no autorizado y evitar contingencias que den lugar a filtraciones; y f) notificar los incidentes tanto a los interesados como a la autoridad nacional de protección.

Además, los partidos, federaciones, coaliciones y candidatos están obligados a llevar un registro de las operaciones de tratamiento de datos realizadas, detallando la información mínima que debe consignarse (artículo 33-C)[102]. El apartado 2 de este artículo establece el deber de conservar los registros de las operaciones durante todo el período electoral, y la obligación se mantiene en caso de que se interponga una demanda para investigar cualquier irregularidad o ilegalidad en el tratamiento de datos por parte de las campañas. La autoridad puede ordenar la presentación de los registros de operaciones, debidamente acompañados de los documentos justificativos (§ 3).

Otro punto interesante que refuerza la protección de datos en el contexto de las campañas es la prohibición de la venta, donación y cesión de datos de clientes por parte de personas jurídicas. Así lo hace el artículo 31, que también prohíbe la venta de registros electrónicos de direcciones y bases de datos de personas físicas o jurídicas. La única excepción, introducida por el párrafo 1-B del artículo 31, se refiere a los datos personales de contacto obtenidos legítimamente por una persona física. En estos casos, se permite su cesión gratuita a los partidos políticos y su utilización cuando

[102] A los efectos de la presente Resolución, los partidos políticos, federaciones, coaliciones, candidaturas y candidatos deben mantener un registro de las operaciones de tratamiento de datos personales, que contenga, como mínimo: I – el tipo de datos y su origen; II – las categorías de interesados; III – la descripción del proceso y la finalidad; IV – la base jurídica; V – la duración prevista del tratamiento, en los términos de la Ley 13.709/2018; VI – el período durante el cual se almacenarán los datos personales; VII – una descripción del flujo de intercambio de datos personales, si corresponde; VIII – los instrumentos contractuales que especifican el papel y las responsabilidades de los controladores y operadores; IX – las medidas de seguridad utilizadas, incluidas las buenas políticas de gobernanza.

exista consentimiento previo de los destinatarios de la publicidad electoral (dueños de los datos)[103].

Además de las normas específicas, las soluciones de IA, para ser utilizadas legalmente en la búsqueda de votos, también están sujetas a la observancia de las normas de campaña de aplicación general. En todos los casos, los candidatos están obligados a registrar sus gastos en los procesos de rendición de cuentas, y también es deseable adoptar buenas prácticas en materia de transparencia y buena fe, con indicación y explicación pública de las herramientas adoptadas y sus respectivos fines.

4.3. Una respuesta global, integral y necesaria

El diseño de la tecnología es también el diseño de la sociedad. Dentro de este diseño tecnosocial, la IA está cobrando protagonismo creciente abriendo un campo de oportunidades en áreas como la investigación científica o los negocios. También puede ayudar a las sociedades a organizarse de manera más eficiente y brindar un mejor servicio a sus ciudadanos, sin descartar situaciones no deseables donde la regla de la ley sea reemplazada por la regla del algoritmo; amenazando los derechos humanos y las libertades fundamentales; y debilitando los procesos democráticos.

El uso de la IA en los procesos electorales se sitúa en un contexto general en el que hemos pasado de la visión utópica de poner en la tecnología todas las esperanzas de regeneración democrática (por su impulso a la transparencia, la participación y la rendición de cuentas) a una visión apocalíptica, según la cual es de la tecnología de dónde vienen todos los males que aquejan a una democracia, que estaría viviendo sus últimos días como consecuencia de la misma. En este contexto, los procesos electorales ocupan un

[103] «§ 1-B. El registro de datos personales de contacto, poseído de forma legítima por una persona natural, podrá ser cedido gratuitamente a un partido político, federación, coalición, candidata o candidato, condicionándose el uso lícito en campaña al consentimiento previo, expreso e informado de los destinatarios en el primer contacto por mensaje u otro medio».

lugar central en la legitimidad del sistema democrático, y el uso de la IA, por su complejidad y opacidad, puede contribuir en esta reacción de rechazo.

La automatización de algunos procesos, a través de los algoritmos, y especialmente a través de sistemas de inteligencia artificial en la maquinaria electoral está transformando la dinámica de las elecciones. De esta manera se han incorporado a las campañas distintos mecanismos que, impulsados por IA, están aumentando el alcance y la eficiencia del proselitismo electoral.

Este tipo de campañas afecta a la «arquitectura de decisión» del votante, propiciando una opinión cada vez más personalizada basada en segmentación y microsegmentación; la generación de contenido con IAG que desdibuja la distinción entre no ficción y ficción, especialmente con el uso de *deepfakes*; y la difusión y redifusión de éstos, a través de mecanismos automatizados. Así lo señaló, Sam Altman, director ejecutivo de OpenAI, en su comparecencia ante el Senado de los Estados Unidos en mayo de 2023 cuando preguntado si pensaba que se podían utilizar los modelos de lenguaje basados en IA, como ChatGPT, para hacer que los votantes se comportaran de determinada manera, mostró su preocupación de que estos mecanismos puedan ser utilizados para persuadir, involucrar y manipular en las relaciones de las candidaturas con los votantes.

A pesar de estas advertencias, la IA tiende a consolidarse como una herramienta indispensable para el ejercicio efectivo de los derechos políticos pasivos, ya que infrautilizar su potencial disminuirá las posibilidades de éxito en las contiendas electorales. En un futuro muy próximo, las campañas sin tecnología, especialmente en las grandes circunscripciones, serán cada vez más caras y menos eficaces, ya que la comunicación generalista parece cada vez menos capaz de movilizar a los votantes. Despreciar el petróleo de datos y herramientas de alto rendimiento que se abre en el horizonte equivaldrá a desafiar hordas de francotiradores a ciegas y sin armas.

El modelo emergente afecta a aspectos centrales de la legitimidad democrática. Parece incuestionable que «los problemas de manipulación de la información son sistémicos» y necesitan «ser

abordados reconociendo el papel de los actores implicados y sus responsabilidades en el mantenimiento de los derechos de los titulares de los datos y de la propia democracia» (Frazão, 2022, p. 567). Este escenario se agrava cuando el arsenal de deconstrucción de la verdad fagocita el núcleo de la agenda política, asumiendo ataques de segmentos *aceleracionistas*[104] que pretenden desestabilizar la sociedad, creando las condiciones perfectas para levantamientos autoritarios que utilizan como pretexto fraudes inexistentes en los procesos electorales.

Los sistemas de IA reconfiguran la esfera pública, disminuyendo su racionalidad y su naturaleza libre e informada. Resucitan deformaciones y desigualdades, afectando al modo en que la información circula y es accedida por el cuerpo social.

Fung y Lessig han puesto un nombre a este fenómeno: Clogger. El punto de partida es que lo más determinante en el «nuevo orden algorítmico», del que forma parte la IA, es que condiciona de manera determinante nuestro acceso a la información, predeterminándola selectivamente. Además de poder manipular los algoritmos de los motores de búsqueda para que los sitios web o las noticias que contengan información falsa aparezcan primero, la IA supondría, como hemos visto, una escalada gradual en la eficacia de los mecanismos de microsegmentación e influencia en el comportamiento, convirtiendo en una realidad asequible el sueño de la personalización del mensaje, gracias a su condición personalizada, dinámica y adaptativa: 1) la generación, gracias a sus modelos de lenguaje, de mensajes en distintos formatos realmente personalizados sin un límite real en cuanto que todo el proceso de segmentación, creación y distribución del mensaje estará automatizado; 2) el uso de técnicas de aprendizaje reforzado, basadas en el *machine learning* y la prueba y error, sirve para generar una serie de mensajes que, adaptándose a las respuestas, van aumentando progresivamente su nivel de persuasión y la eficacia de los

[104] Dentro de la literatura sobre movimientos extremistas, el «aceleracionismo» caracteriza a los grupos que «pretenden acelerar el caos», utilizando para ello diversos medios, entre ellos el uso de redes digitales para expandir y consolidar el alcance de diversas teorías conspirativas (Prado, 2023, p. 226).

mismos; 3) este aprendizaje se alimentará también de la respuesta de los demás votantes, a través de conversaciones dinámicas, en las que cada vez se afinan más los mensajes que están ofreciendo mejores resultados persuasivos en comunidades similares en un momento concreto, utilizándoles para el resto de la comunidad, con la consiguiente adaptación a los elementos particulares de cada uno. Además, apuntan también los autores, esta labor de persuasión podrá desarrollarse incluso sin utilizar mensajes estrictamente políticos, robando la atención de votantes seleccionados por su simpatía con el adversario político con mensajes sobre sus aficiones, mucho más atractivas que la política, o de mensajes o anuncios desagradables, que «oculten» los contenidos políticos. O, como también hemos visto, a través de la persuasión a los grupos de confianza del votante ajeno a la política al que se intentaría persuadir a través de microcomunidades autoreferenciales, silos de verdades distintas compartidas solo por iguales. Sin que los afectados tuvieran forma de saber las estrategias utilizadas en este ejercicio de persuasión (Fung; Lessig, 2023).

El uso de estas técnicas por parte de uno de los candidatos, como cualquier tecnología que irrumpe en campaña, generaría un efecto imitación, que provocaría en un breve espacio de tiempo su generalización y un riesgo de convertir la campaña en una guerra entre tecnologías, en la que triunfaría la más efectiva, lo que supondría en la práctica el fin de la democracia, pese a la apariencia de la misma, al mantenerse los elementos propios de un proceso electoral como discursos, anuncios, mensajes, votación y recuento pero en un espacio donde la opinión pública además de fragmentada sería inducida, irreal y falsificada.

Sin embargo, el riesgo más grave para la democracia parece ser la erosión de la confianza: en la era de la IA, el consenso fiduciario se ve socavado por falacias de alto rendimiento, y el tejido social se deshilacha en un entorno de sospecha permanente. La desconfianza y la división se unen para elevar las hostilidades, la inseguridad y la polarización, revelando un mundo en el que la inestabilidad trata de imponerse como la nueva normalidad.

El odio político y la polarización radical envenenan el proyecto democrático, que está ligado a la condición del pluralismo. La

vigencia de los ideales de respeto, tolerancia y diálogo sería el corolario de un modelo de sociedad en el que las diferencias no se hacen cada vez mayores, sino que interactúan en la búsqueda de posibles consensos (Frazão, 2022, p. 567). Las elecciones normales y legítimas sólo prosperan cuando prevalecen los valores democráticos, lo que exige —de la sociedad en su conjunto— una mayor atención a las externalidades de la normalización del uso de estos medios tecnológicos desde una perspectiva social.

El desarrollo de la democracia digital —con la IA en su núcleo— cambia el comportamiento social, dicta un nuevo régimen de información política y reconfigura las condiciones competitivas de la arena electoral. Implica un nuevo conjunto de retos sistémicos para las instituciones electorales y el sustrato de la ciudadanía y, en consecuencia, señala la necesidad de revisar los acuerdos, con la vista puesta en salvaguardar los supuestos de libertad, igualdad, integridad, transparencia y justicia. Tanto comprender y desmitificar las amenazas, como identificar y abordar las vulnerabilidades son pasos esenciales para mantener la soberanía popular a salvo del fraude, la trampa y la manipulación.

4.3.1. Las bases de la regulación

Ante la ausencia de una regulación específica para las elecciones es necesario establecer un marco legal claro y sólido sobre el uso de la IA en campaña, sin delegarlo en las empresas del sector. Un marco que adopte una posición sobre: a) la protección de la privacidad; b) la responsabilidad del contenido producido con IA; c) los modelos de publicidad donde la IA desempeña un papel fundamental, para evitar que afecte a la libertad de decisión, aumentando la transparencia (con bibliotecas de anuncios políticos de acceso abierto) y posicionándose con claridad sobre el uso del microtargeting estableciendo limitaciones a su utilización o incluso planteando la prohibición al uso de estas técnicas en procesos electorales (sobre la base de que la libertad de expresión no protege el discurso de las máquinas); d) el respeto a las garantías del debido proceso cuando se deje en manos de la IA las medidas de

moderación y eliminación de contenido así como el uso de esta tecnología para luchar contra la desinformación (para garantizar que sea realmente independiente); e) el nivel de transparencia exigible (con medidas como las que evitan que los *bots* se presenten como personas identificando el contenido generado por máquinas); así como f) la rendición de cuentas, facilitando el acceso de investigadores, las iniciativas de verificación de datos y las organizaciones de la sociedad civil para evaluar el impacto de la IA en las campañas políticas en línea.

Para hacerlo los legisladores deberán responder una serie de preguntas clave: cuál es el propósito de la regulación, qué conductas deberían cubrirse, cómo regular esas conductas y quién debería ser responsable, específicamente si la regulación debería apuntar solo a aquellos que crean o difunden *deepfakes* o también a las plataformas en línea que facilitan su transmisión.

La regulación del uso de IA en campaña es crucial para promover un electorado informado y proteger la integridad del proceso electoral. Por eso debe tratar de dar respuesta a los distintos riesgos democráticos relacionados con el papel de la IA en los procesos electorales, que hemos analizado en nuestro trabajo. En concreto:

1. *En cuanto al mal uso*: contener interferencias ajenas, patologías algorítmicas, mensajes microdirigidos y chantajes psicométricos.
2. *En cuanto a la manipulación de la opinión pública*: hacer frente a la desinformación y las teorías de la conspiración, desacreditar el comportamiento inauténtico[105] y *los deepfakes*.

[105] El término se refiere a «técnicas utilizadas para publicar, promover y divulgar, artificialmente la participación orgánica en plataformas digitales a través de cuentas falsas que interactúan validando comentarios, distribuyendo *likes* o compartiendo publicaciones entre ellas» (Cardiel Soto; Alvim; Rondon, 2022, p. 53). En este contexto, se argumenta la necesidad de nuevos dispositivos para frenar la adquisición de cuentas *bots*, así como la instalación de *clickfarms*, granjas de *trolls*, etc.

3. *Respecto a la opinión autodeterminada*: reforzando la autonomía de los ciudadanos, capacitándolos para evaluar críticamente la comunicación y desarrollando habilidades para distinguir los hechos de las opiniones, y para comprobar la consistencia de las pruebas y las líneas argumentales (educación informativa); evitar la exposición a informaciones falsas, evaluando la credibilidad de las fuentes, identificando las características de las *fake news* y fomentando el uso ético y responsable de las redes sociales (educación mediática)[106]; difundir los valores constitucionales, basados en la corresponsabilidad entre la clase política y la sociedad civil y en la necesidad de paz, tolerancia y respeto entre todas las personas (educación para la cultura democrática).
4. *En cuanto a las perturbaciones de la información*: aumentar la supervisión de la publicidad en línea, establecer sistemas de etiquetado y verificar la autoría del contenido de las campañas políticas generadas por IA.
5. *En cuanto al uso de sistemas de IA con un sesgo discriminatorio*: regular las plataformas de medios sociales para garantizar una moderación justa y algoritmos neutrales. Los algoritmos se han convertido en el «aparato disciplinario de nuestro tiempo» (Beiguelman, 2021, p. 46) y, como tal, es necesario reformar su comprensibilidad para que *las partes interesadas* puedan tener claridad y estar seguras de su funcionamiento ético, responsable y políticamente neutral.
6. *En cuanto al acceso a la información*: garantizar la neutralidad en línea y una Internet abierta, de modo que cualquier restricción de acceso se base en disposiciones legales. Los

[106] «En una sociedad cada vez más digital no existe un acceso igualitario a la información procedente de las nuevas tecnologías de la información y la comunicación, ocasionando una brecha digital y una desigualdad de oportunidades para los colectivos más vulnerables». En este escenario, surge la necesidad de «nuevas competencias como la selección de la información, la codificación, la búsqueda de la misma y la capacidad de interpretación, para que los individuos estén menos expuestos a la desinformación y la infoxicación» (Morales Romo, 2023, p. 260).

motores de búsqueda y las plataformas, hemos visto que: a) imponen normas computacionales opacas que ocultan selectivamente la pluralidad de puntos de vista y mercantilizan la atención de los usuarios; b) normalizan campañas hiperpersonalizadas que aíslan a las personas y las someten a una dieta informativa arbitraria y restrictiva; c) potencian campañas de desinformación, amplificando discursos de odio y contenidos nocivos; d) desequilibran los procesos electorales, estructurando un campo de juego asimétrico en el que el acceso a la tecnología de datos segrega a los desfavorecidos de los favorecidos. A la luz de estos aspectos, no es exagerado afirmar que «la revolución de las comunicaciones digitales […] ha estado maltratando nuestras elecciones» (Moore, 2022, p. 11).

7. Con respecto *a las ciberamenazas de IA*: ofrecer respuestas rápidas y eficaces, contando con la colaboración de las plataformas, en un equilibrio permanente para que la reducción de la circulación de mensajes no degenere en un modo de censura privada[107].
8. En lo que se refiere a los *sujetos*, la mayoría de las regulaciones deberían dirigirse a candidatos, comités de acción política (PAC) y otros que utilizan estas técnicas poniendo en riesgo la integridad de las elecciones, pero no deben de dejar a un lado regulaciones específicas para las plataformas en línea y otros intermediarios que son indispensables para la creación y divulgación de este tipo de contenidos.

En resumen, es necesaria una regulación equilibrada que aborde los riesgos planteados por el uso de estas técnicas sin restringir excesivamente la libertad de expresión.

---

[107] En este contexto, se necesitan mecanismos de transparencia, responsabilidad y rendición de cuentas en relación con las aplicaciones tecnológicas de IA y los intermediarios de Internet en el contexto de los procesos electorales. También es necesaria la adopción obligatoria de códigos éticos y de responsabilidad social corporativa entre los proveedores de servicios digitales y, en este caso, los desarrolladores de aplicaciones de inteligencia artificial.

4.3.2. El papel central de los organismos electorales

Para alcanzar estos objetivos, es imprescindible capacitar a los organismos electorales para hacer frente a situaciones de emergencia, dotándolos de los recursos necesarios y de la formación específica que permita la integración efectiva de las tecnologías de la información y la comunicación, incluida la IA, así como la gestión de los riesgos asociados a la ciberseguridad. Esto implica reforzar sus competencias e incrementar sus recursos materiales y humanos, de forma que estén preparados para hacer frente a la complejidad que conlleva la creciente automatización, siempre con la conciencia de que, hoy por hoy, no es posible hacer frente a la IA sin IA[108].

La resiliencia democrática y la correspondiente reorganización de los órganos electorales deben desarrollarse paralelamente a la preservación de las libertades individuales y de acuerdo con la comprensión de sus límites, algo fundamental si se quiere hacer frente a las amenazas cognitivas —cada vez mayores y más complejas— de acuerdo con los dictados de la democracia. El mantenimiento de un «entorno virtual sano […] sólo puede realizarse prestando atención a una libertad de expresión libre de tergiversaciones» (Branco; Branco, 2022, p. 67).

---

[108] «Sin embargo, en "el futuro" tanto la "ofensiva" como la "defensiva" —tanto la difusión de la desinformación como los esfuerzos para combatirla— se automatizarán cada vez más y se confiarán a la IA. La IA generadora de lenguaje GPT-3 ha demostrado su capacidad para crear personajes ficticios, utilizarlos para producir un lenguaje característico del discurso del odio y entablar conversaciones con usuarios humanos para inculcarles prejuicios e incluso incitarles a la violencia. Si una IA de este tipo se desplegara para difundir el odio y la división a gran escala, los seres humanos por sí solos podrían no ser capaces de combatirla el resultado. A menos que la IA se detenga en una fase temprana de su despliegue, identificar y desactivar manualmente todo su contenido mediante investigaciones y decisiones individuales resultaría muy difícil, incluso para los gobiernos y los operadores de plataformas de redes más sofisticados. Para una tarea tan vasta y ardua, tendrían que recurrir —como ya lo hacen— a algoritmos de moderación de contenidos de IA. Pero, ¿quién y cómo los crea y supervisa?» (Kissinger; Schmidt; Huttenlocher, 2023, p. 105).

La IA introduce nuevas herramientas que cambian las reglas del juego democrático. Al modificar la esfera comunicativa, la IA está desafiando las estructuras tradicionales de la sociedad y transformando la lógica contemporánea de las competiciones electorales, como ha ocurrido otras veces a lo largo de la historia. En este contexto, las organizaciones electorales se enfrentan a un doble reto: a) incorporar estas tecnologías para ampliar y proteger los principios democráticos y b) mitigar los efectos adversos de las mismas sobre la integridad de los procesos electorales, y hacerlo de manera eficaz.

Pero la respuesta, que como hemos visto resulta más necesaria que nunca, debe ir mucho más allá de lo jurídico, y necesita además para ser eficaz de componentes tecnológicos, comunicativos, culturales, políticos y educativos. Esta visión integral que debe tener una perspectiva global y respetuosa de los principios democráticos, garantizando no sólo la libertad de expresión y el derecho a la información sino también la privacidad y la libertad del voto, el derecho a participar en los asuntos públicos, la equidad en los procesos o la integridad de los mismos, principios todos que se ven afectados por el uso de la IA en campaña.

La garantía de la libertad de expresión informativa, si bien «protege la emisión de opiniones, convicciones, comentarios, valoraciones o juicios sobre cualquier materia, incluidas las cuestiones de interés público», puede —a pesar de su posición preferente— sufrir restricciones[109], siempre que estas estén establecidas en la ley (principio de legalidad), que esta busque una finalidad legítima, y que la restricción evite un daño mayor que el que cau-

[109] «Desde esta perspectiva, los derechos fundamentales de acceso a la información y de libertad de expresión deben ser interpretados de la mejor manera posible, a fin de garantizar la preservación de sus núcleos esenciales. Y es que las informaciones falsas pueden debilitar estos derechos —tan caros a la construcción del Estado Democrático de Derecho— en la medida en que se difundan sin ningún análisis sobre la veracidad de su contenido. Por esta razón, las *fake news* influyen en la formación de la opinión pública de tal manera que ponen en peligro la formación del Estado Democrático de Derecho, ya que eliminan la capacidad del ciudadano de seguir los actos de la vida pública de forma íntegra y coherente» (Mendes, 2022, p. 75).

sa (necesidad y proporcionalidad). La libertad de expresión está directamente relacionada con el derecho de acceso a una información adecuada y es la combinación de ambas la que establece los criterios de proporcionalidad, que en el ámbito electoral hay que juzgar con especial atención. Pero la libertad de expresión no es un obstáculo para el derecho a la información sino su principal garantía. De ahí que, en período electoral, defender la libertad de expresión sea el único camino para «proteger el acceso de los ciudadanos a una información veraz, para que puedan tomar sus decisiones personales, sociales o políticas de manera fundada» (Mendes, 2022, pp. 70 y 73)[110], tratar de contraponer la defensa del derecho a la información frente a la libertad de expresión sería como cortar el oxígeno para evitar que se propagaran los virus, una garantía de muerte por asfixia.

Salvaguardar la integridad de las campañas significa proteger los procesos electorales y la democracia. Para ello, es importante entender los profundos cambios que las nuevas tecnologías están provocando para diseñar respuestas proporcionadas y contundentes que impidan que el debate público descienda a una suerte de parloteo artificial, en el que se distorsiona la realidad, donde las

[110] «La imagen de un mercado de ideas, en el que el choque entre ellas hará prevalecer la más positiva y emerger la verdad, es importante en el sentido de reconocer la libertad de expresión en sentido amplio como elemento central de la democracia y de la autonomía individual. Sin embargo, su adopción acrítica, sin limitaciones a determinados discursos, representa un riesgo para la propia democracia. Un modelo de libre mercado sin restricciones hace posible que informaciones alejadas de la finalidad de la divulgación y que ponen en cuestión el sistema democrático actúen de forma ilimitada e incluso prevalezcan, ante la evidente constatación de que las personas no siempre son capaces de identificar racionalmente la falsedad y la manipulación de la verdad» (Biolcati, 2022, p. 127). Adaptando la gráfica explicación de Rafael Alcácer sobre el discurso del odio, en términos de interpretación constitucional la desinformación es vista como un elemento delimitador negativo del derecho a la libertad de expresión, de modo que la relación entre contenidos desinformativos y libertad de expresión será la de «círculos tangentes»: lo que desinforma no entra dentro de la libertad de expresión, y viceversa (Alcácer, 2023, p. 41), añadiendo que el mismo razonamiento se aplica a los ataques a grupos vulnerables o minoritarios (Piras, 2021, p. 32).

intimidades son *hackeadas* para ejercer presiones indebidas y en el que el algoritmo cierre las puertas a la pluralidad del mundo. En definitiva, para garantizar que las tecnologías de la IA se apliquen en términos compatibles con la equidad y la autenticidad, con la autonomía y la libertad para definir el voto.