ANÁLISIS

DEMOCRACIA Y DERECHOS HUMANOS

REGULACIÓN PARA COMBATIR LA DESINFORMACIÓN.

ESTUDIO DE OCHO CASOS INTERNACIONALES Y RECOMENDACIONES PARA UN ENFOQUE DEMOCRÁTICO

João Brant / João Guilherme Bastos dos Santos
Tatiana Dourado / Marina Pita

Junio 2020

El uso de información falsa y engañosa para cambiar el comportamiento de nichos específicos de la sociedad es una práctica que existe desde mucho antes de su popularización en Internet.

El análisis revela los desafíos al proceso de construcción de soluciones reguladoras que protejan derechos y sean efectivas para promover el acceso a información confiable.

La desinformación está profundamente entrelazada con los procesos políticos y sociales, y no es posible pensar en soluciones legales y reguladoras aisladamente del entendimiento de estos procesos y de los contextos nacionales y locales.

FRIEDRICH EBERT STIFTUNG

DEMOCRACIA Y DERECHOS HUMANOS

REGULACIÓN PARA COMBATIR LA DESINFORMACIÓN.

ESTUDIO DE OCHO CASOS INTERNACIONALES Y RECOMENDACIONES PARA UN ENFOQUE DEMOCRÁTICO

CONTENIDO

1

INTRODUCCIÓN

El uso estratégico de información falsa y engañosa buscando alterar el comportamiento de nichos específicos de la sociedad es una práctica que existe desde mucho antes de su popularización en Internet. Desde el uso de propaganda política en las guerras (Lasswell, 1938) hasta la adopción de atajos de información, donde los votantes llenan los vacíos de información sobre política (Popkin, 1994), la difusión deliberada de rumores y falsificaciones se adecua a los medios de comunicación y las infraestructuras tecnológicas imperantes en cada momento histórico. Ya sea para buscar soldados dispuestos a abandonar el campo de batalla o a un grupo demográfico más dispuesto a abstenerse, rechazar o votar por un candidato si están en contacto con información específica, el efecto de estas campañas depende de la efectividad de la identificación de nichos y la focalización / entrega de información personalizada a ellos.

En la última década, el aumento de la fragmentación y opacidad de los entornos informativos organizados en plataformas digitales como Facebook y WhatsApp ha generado el 'entierro' de parte del debate público[1]. Esta característica dada por la profusión de grupos cerrados impide el escrutinio público sobre parte de las ideas en discusión y dificulta la visibilidad de perspectivas contradictorias.

Sin transparencia, la confiabilidad de la información se debilita. En espacios de discusión invisibles para el público y sin responsabilidad moral y legal de los interlocutores, es más probable que prospere la información engañosa y fraudulenta. Con todos los límites del modelo de comunicación tradicional y su dificultad para brindar un debate plural y diverso, la transparencia del debate público y la búsqueda de credibilidad por parte de los medios contribuyeron a hacer de la confiabilidad de la información un problema menos relevante hasta la década de 2010.

La práctica de la desinformación ha sido definida por la Unesco (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura) y por el PNUD (Programa de las Naciones Unidas para el Desarrollo) como "contenido falso, manipulado o engañoso, creado y diseminado intencionalmente o no, y que puede causar daños potenciales a la paz, a los derechos humanos y al desarrollo sostenible". De hecho, sus efectos se sienten en muchos campos: en campañas electorales, en asuntos de salud pública (como fue evidente en la pandemia covid-19), en la difusión de discursos de odio contra grupos sociales o en el ataque a la reputación de activistas, y en todas las disputas relevantes en el campo socioambiental, por mencionar solo los ejemplos más evidentes.

La tendencia a la segmentación del perfil de audiencia ayuda a comprender la inversión de grupos de interés, gobiernos y campañas políticas en el cruce de datos que garantiza una mayor precisión en la entrega y persuasión al efecto del mensaje. Antes de la popularización de plataformas y aplicaciones, sin embargo, los datos involucrados en el tratamiento de perfiles eran de difícil acceso, como en el cruce de datos como tarjetas de crédito, revistas favoritas y lugar de residencia (Howard, 2006). La entrega de información supuso otro conjunto

1 Ver Brant, João. Modelo de aplicativos de mensagens enterra o debate público. *Folha de São Paulo*, 1 de noviembre de 2020. Disponible en: <https://www1.folha.uol.com.br/ilustrissima/2020/10/modelo-de-aplicativos-de-mensagens-enterra-o-debate-publico.shtml>

de esfuerzos, en servicios distintos a los que intervienen en el tratamiento de los datos del perfil.

Las plataformas y aplicaciones basadas en perfiles personales y la entrega de publicidad dirigida hacen que sea mucho menos costoso obtener y verificar datos personales (incluidos los relacionados con la posición política) y también unifican la fuente de los datos y la entrega de información dirigida. Las campañas de desinformación se ajustan a este tipo de escenario porque dependen de la identificación de nichos y la entrega personalizada de mensajes y anuncios políticos. Herramientas como *dark post* (donde solo un tipo de audiencia puede ver la información publicada, sin que sea visible para todos en la línea de tiempo de la página de la campaña), y la propia segmentación de la audiencia en cajas de resonancia para reafirmar creencias y valores, reduce los posibles efectos secundarios que estas publicaciones podrían tener en otras audiencias.

En 2016, dos eventos altamente polarizados, las elecciones en los Estados Unidos y el referéndum del Brexit, revelaron nichos que probablemente cambiaron su comportamiento con base en campañas de desinformación. La reacción política puso en el centro de la discusión la viabilidad y la necesidad de regular las plataformas *online* para evitar su uso en campañas basadas en información falsa y engañosa. En Brasil, las elecciones presidenciales de 2018 también estuvieron marcadas por prácticas de desinformación con un impacto relevante en la percepción de los votantes[2].

En pocos años, este asunto se convirtió en una agenda política central en el mantenimiento de los procesos democráticos contemporáneos, lo que ha llevado a la convocatoria de miembros y propietarios de empresas tecnológicas a declarar en diferentes parlamentos, al desarrollo de leyes nacionales para tratar el problema y a la experimentación de regulaciones por parte de las propias empresas.

Las audiencias con miembros de Cambridge Analytica en el Parlamento británico y en el Congreso estadounidense con propietarios de empresas como Facebook, así como los grupos de trabajo de la Unión Europea, marcan el inicio de la popularización de este debate, que ganó aún más visibilidad con películas como *Privacidad hackeada* (*The Great Hack*/2019) y *El dilema social* (*The Social Dilemma*/2020). En Brasil, la CPMI (Comissão Parlamentar Mista de Inquirido) de las *fake news* recogió denuncias de disidentes del bolsonarismo sobre el funcionamiento de la denominada Oficina del Odio, además de otros expertos en la materia.

En este sentido, la producción descentralizada de publicaciones que circulan en estas plataformas digitales genera desafíos transnacionales de difícil solución. Por un lado, los usuarios y las campañas políticas pueden apropiarse de las herramientas de denuncia para atacar a los oponentes. Por otro, ante la imposibilidad de analizar miles de posts y horas de videos en un solo día, las empresas invierten en filtros algorítmicos que actúan para identificar contenidos y perfiles potencialmente falsos. Sin embargo, al filtrar y tomar decisiones que implican cuestiones controvertidas, este tipo de herramienta traslada las decisiones sobre las publicaciones consideradas posiblemente perjudiciales a los programadores y al código de aprendizaje automático, a menudo sin revisión humana. Los sesgos en los motores de búsqueda y el análisis de imágenes, que reproducen patrones racistas y sexistas a partir del aprendizaje automático, son quizás el ejemplo más conocido de problemas en este sentido.

Uno de los desafíos del fenómeno de la desinformación es que, dado que se propaga significativamente en espacios 'subterráneos', no puede medirse adecuadamente. Investigadores de todo el mundo se dedican a fortalecer los métodos y las condiciones de investigación, pero

2 Véase Santos, João Guilherme Bastos y Freitas, Miguel. WhatsApp, política móvil y desinformación: ¿cómo se dio la viralización de las noticias falsas en las elecciones brasileñas? Centro de Estudios en Libertad de Expresión y Acceso a la Información. Universidad de Palermo: Buenos Aires, 2019. Disponible: https://www.palermo.edu/Archivos_content/2020/cele/febrero/WhatsApp-politica-movil-y-desinformacion.pdf; y Solano, Esther; Brant, João; Brito Cruz, Francisco et al. Secretos y mentiras: WhatsApp y las redes sociales en las elecciones presidenciales de Brasil en 2018. Centro de Estudios en Libertad de Expresión y Acceso a la Información. Universidad de Palermo: Buenos Aires, 2019. Disponible en: <https://www.palermo.edu/Archivos_content/2020/cele/febrero/Secretos-y-mentiras-WhatsApp-y-las-redes-sociales%20.pdf>

han estado trabajando con muestras limitadas, que no son una muestra de un universo complejo e indefinible. Además, hay poca colaboración de las empresas para dar transparencia a sus prácticas y procesos de intercambio de contenidos. Esto hace que sea difícil comprender también las tendencias de crecimiento o disminución en la ocurrencia del fenómeno.

Al mismo tiempo, el fenómeno de la desinformación está profundamente entrelazado con los procesos políticos y sociales, y no es posible pensar en soluciones legales y reguladoras aisladamente del entendimiento de estos procesos y de los contextos nacionales y locales.

Además, la legislación que enfrenta el problema de la manipulación de la información enfrenta desafíos multiplataforma y transnacionales. Esto se debe a que la información viaja a través de diferentes plataformas interconectadas por medio de herramientas de intercambio, con diferente lógica de filtrado y términos de uso, y los países por donde circula esta información no son los mismos donde se ubican las empresas y sus bases de datos. Los esfuerzos coordinados de empresas sin este tipo de incentivo aún son limitados en diferentes países.

Este trabajo analiza casos en los que diferentes países (Alemania, Francia, Canadá, India y Singapur) buscaron crear reglas y conjuntos de prácticas para mitigar los efectos nocivos de las campañas *online* centradas en información falsa o engañosa. También se analiza una iniciativa regional, el Código de prácticas de la Unión Europea. Se presentan también iniciativas presentadas, pero aún no aprobadas, en el Reino Unido y Brasil.

2

ESTUDIO DE CASOS

2.1 ALEMANIA NETZDG

La principal referencia para la regulación del discurso en las plataformas en línea es la Network Enforcement Act (Netzwerkdurchsetzungsgesetz - NetzDG) de Alemania, cuyo objetivo es contener el discurso de odio y otros contenidos y expresiones tipificados como delito por el Código penal del país. Para lograr el objetivo, la NetzDG, aprobada en septiembre de 2017, creó una serie de obligaciones para las empresas de redes sociales que tienen un objetivo de lucro y más de dos millones de usuarios registrados en Alemania.

Debido a la historia de crímenes de lesa humanidad perpetrados en el país, la regulación de las restricciones al discurso de odio y, o cualquier manifestación de apoyo al nazismo u otras organizaciones consideradas inconstitucionales es robusta, lo que no impide que Alemania se encuentre entre los diez países que más respetan la libertad de expresión, en una encuesta de la organización Artículo 19 para los años 2019/2020 (Artículo 19, 2020a) y sea también el undécimo país que más respeta la libertad de prensa en la lista de Reporteros sin Fronteras.

La legislación se puede comparar con la propuesta del Reino Unido de crear un deber de cuidado, con parámetros, pero sin la responsabilidad de las plataformas por el contenido individual publicado por terceros. La ley obliga a las empresas a retirar los contenidos ilegales o las conductas delictivas previstas en el Código penal –y a remitir las denuncias a los organismos competentes–; a elaborar informes de transparencia sobre la gestión de los contenidos y las denuncias de los usuarios; y a garantizar mecanismos de recurso y seguimiento de las medidas adoptadas a los denunciantes. Además, deben capacitar a los equipos de moderación en la legislación alemana y designar a un representante legal en Alemania para responder a las solicitudes de regulación. La ley también crea un órgano administrativo para aplicar sanciones y la posibilidad de una entidad autorreguladora certificada por el órgano administrativo.

La disposición más controvertida aprobada es la creación de un plazo ajustado, veinticuatro horas, para la eliminación de contenido "evidentemente ilegal", expresión considerada amplia por la mayoría de los críticos, que podría llevar a la sobreintervención de plataformas que priorizarían la seguridad financiera y legal. Existe la posibilidad de una reacción en un periodo de tiempo más largo, si es necesario un análisis detallado para clasificar el contenido. El periodo de siete días se puede extender si la empresa requiere el apoyo de la entidad autorreguladora certificada por la entidad administrativa gubernamental.

Las redes sociales no pueden ser sancionadas por errores individuales en la gestión del contenido denunciado. Solo se pueden imponer multas por violaciones "sistemáticas" de la ley. Es decir, si la red social comete un error de apreciación o es negligente o gestiona un elemento de forma errónea, no tendrá responsabilidad. Y, sin embargo, el valor de las multas por fallas sistémicas de aplicación de la ley puede alcanzar los cincuenta millones de euros. En junio de 2019, Facebook fue multada con dos millones de euros por no informar el número de quejas que recibió sobre contenido ilegal en su plataforma. A diferencia de Alphabet, que controla YouTube y Google, y de Twitter, Facebook no ha realizado ninguna adaptación a la plataforma para permitir que los usuarios denuncien fácilmente contenido ilegal de acuerdo con las disposiciones de la ley. La opción de

la empresa fundada por Mark Zuckerberg fue crear un formulario de reclamaciones en páginas no fácilmente accesibles, lo que implicaba un menor cuestionamiento de contenidos[3]. Además, hay afirmaciones de que Facebook está eligiendo clasificar el contenido como una violación de las propias reglas de la empresa, en lugar de infracciones legales.

Vale la pena mencionar que gran parte de la controversia que involucra a NetzDG se apaciguó al excluir la disposición para controlar nuevas cargas de contenido prohibido, ya que el contexto o los comentarios en torno a una imagen o video pueden ser determinantes para configurar el discurso ilegal o no. Sin embargo, organizaciones dedicadas a frenar el terrorismo señalan que la ausencia del dispositivo ha hecho que la ley sea menos eficaz y eficiente.

En el momento de la discusión del proyecto de ley, la propuesta contaba con la oposición de importantes organizaciones[4]. El relator de las Naciones Unidas para la libertad de expresión, David Kaye, expresó su preocupación por el posible impacto de elevar el nivel de gestión de contenido de terceros por parte de plataformas digitales, por temor a ser responsabilizado por fuertes sanciones, lo que podría significar censura privada. Los primeros análisis del impacto de la ley muestran que las plataformas no optaron por retirar los contenidos tan pronto como fueron denunciados, ya que menos del 20% de los elementos señalados por los usuarios como ilegales fueron objeto de retirada[5]. A la misma conclusión llegó el estudio realizado por el profesor de la Universidad de Humboldt en Berlín, Martin Eifert, que concluyó que no había un efecto de bloqueo excesivo y que la ley era eficaz[6]. Sin embargo, los críticos señalan que no hay pruebas de que se haya reducido la circulación de la incitación al odio y otras expresiones prohibidas[7].

A pesar de las críticas de entidades especializadas en libertad de expresión, NetzDG cuenta con un fuerte apoyo popular (Jacob, 2018). Recientemente, incluso se aprobaron nuevas disposiciones, a saber: las empresas sometidas a NetzDG deberán informar de manera proactiva los casos graves de discurso de odio a las autorida-

3 Después del plazo para la publicación del primer reporte de transparencia bajo el Network Enforcement Act en el verano de 2018, los medios de comunicación alemanes relataron que Facebook presentó un número mucho más bajo de reclamos que otras redes sociales. El reporte de Facebook listó 886 reclamos sobre contenido ilegal en 1.704 posteos, de los cuales 362 fueron excluidos. Ese número fue significativamente más bajo que los presentados por Google y Twitter. Google recibió 215.000 reclamos por posteos en su plataforma de video, YouTube, y excluyó a 58.000 posteos, mientras Twitter reportó 265.000 reclamos y removió 29.000 posteos (Bundesamt fur Justiz, 2019).

4 Entre ellas Open Knowledge Foundation, Chaos Computer Club, Reporteros sin Fronteras Alemania y Wikimedia Foundation Alemania. Artigo 19 también se manifestó contrario a la propuesta.

5 "This study shows that the reality is in between these extremes. NetzDG has not provoked mass requests for takedowns. Nor has it forced internet platforms to adopt a 'take down, ask later' approach. Removal rates among the big three platforms ranged from 21.2% for Facebook to only 10.8% for Twitter" (Ethikson, 2018). (Traducción nuestra: "Ese estudio muestra que la realidad está entre esos extremos. NtzDG no ha provocado solicitudes masivas de retiros. Tampoco ha obligado a las plataformas de Internet a adoptar un enfoque de 'retirar primero, preguntar después'. Las tasas de retiro entre las tres grandes plataformas oscilaron entre el 21,2% para Facebook y solo el 10,8% para Twitter" (Ethikson, 2018).

6 Cf. Campos, 2020.

7 "Before delving into the transparency reports, it is important to note that these data only cover removal decisions arising from NetzDG complaints, and do not account for other removals based on other types of complaints, referrals, or injunctions. Furthermore, the metric of takedowns does not reveal whether NetzDG has achieved its purpose of combating hate speech and other online excesses. The differences between complaint mechanisms and the reports themselves make certain types of comparison difficult. It is also hard to know how the volume of content removal compares to the overall volume of illegal speech online" (Tworek, 2019). (Traducción nuestra: "Antes de profundizar en los informes de transparencia, es importante tener en cuenta que estos datos solo cubren las decisiones de eliminación que surgen de las quejas de NetzDG y no tienen en cuenta otras eliminaciones basadas en otros tipos de quejas, referencias o mandatos judiciales. Además, la métrica de eliminaciones no revela si NetzDG ha logrado su propósito de combatir el discurso de odio y otros excesos en línea. Las diferencias entre los mecanismos de denuncia y los mismos informes dificultan ciertos tipos de comparación. También es difícil saber cómo se compara el volumen de eliminación de contenido con el volumen general de habla ilegal en línea" (Tworek, 2019).

des, lo que ha llamado la atención de los defensores de la privacidad, quienes señalan que la norma violaría los derechos previstos por la ley.

2.2 FRANCIA

Francia ha continuado su tradición reguladora y es un exponente en cuanto a los esfuerzos para regular la expresión ilegal en línea. En diciembre de 2018, el parlamento aprobó la ley contra la manipulación de la información (ley 1202/2018), que tiene como objetivo evitar la injerencia extranjera en las elecciones y aumentar la transparencia en los anuncios en plataformas digitales durante el periodo electoral.

La ley establece que los proveedores de plataformas digitales tienen el deber de cooperar en la lucha contra la desinformación; la obligación de designar a un representante legal que sea su punto de contacto en el territorio francés; crear un medio visible y de fácil acceso para que los usuarios señalen la información falsa y presentar una declaración anual al Conseil Supérieur de l'Audiovisuel (CSA), el regulador de las comunicaciones, detallando las medidas adoptadas contra la difusión de información falsa. Francia ya contaba con una normativa que tipificaba como delito la difusión de desinformación. Las plataformas también tienen el deber, durante el periodo electoral (definido como tres meses antes del primer día de las elecciones generales hasta la votación), de garantizar la transparencia en relación con el contenido de la información patrocinada vinculada a los debates de interés público, indicando la identidad, el valor y el modo en que se utilizan los datos personales.

La ley también creó un nuevo procedimiento judicial para establecer la interrupción de la difusión de una alegación o imputación inexacta o engañosa de un hecho que pueda alterar deliberadamente la imparcialidad de la próxima votación o que se esté difundiendo artificial o masivamente por medio de un servicio de comunicación en línea. Los casos deben ser analizados por un juez que debe emitir una decisión dentro de las cuarenta y ocho horas. La ley refuerza aún más el poder del CSA para contener cualquier intento de una campaña de desestabilización o desinformación por parte de un servicio de televisión controlado o influenciado por un estado extranjero.

Pese a la polémica, la ley, que no prevé sanciones en las redes sociales, fue considerada constitucional con ajustes mínimos. Las principales críticas apuntan a las amplias definiciones de contenido que se pueden suprimir, lo que permite un análisis muy subjetivo de plataformas y jueces. Además, existe una gran preocupación de que la ley sea explotada principalmente por políticos y grandes empresas, con mayor capacidad de litigio. La ley hace recaer una mayor responsabilidad en las plataformas de Internet, cuya interpretación del alcance de los contenidos a retirar puede ser mucho más amplia que las sentencias anteriores del Conseil constitutionnel (Tribunal Constitucional francés) y del Tribunal de Justicia de las Comunidades Europeas.

Sin embargo, la aprobación de la ley en 2018 no fue suficiente para responder a todas las demandas de regular el discurso en línea. En mayo de 2020, el parlamento francés aprobó la ley de lucha contra el odio en línea, que busca responder a la creciente incidencia de discursos de odio e incitación a la violencia en línea. La denominada "ley Avia" fue criticada por grupos de diferente espectro político, organizaciones de la sociedad civil como Artículo 19 y la Electronic Frontier Foundation, además de la Comisión Europea (European Commission, 2019), entre otros.

La ley, de la que posteriormente el Tribunal Constitucional consideró como inconstitucional la mayoría de sus artículos, establecía la obligación de todos los sitios de eliminar el contenido que implicara abuso sexual infantil y terrorismo, previa notificación de la policía y otros órganos de la administración pública, dentro de una hora después de la notificación.

Además, la ley requería que las redes sociales y los motores de búsqueda evaluaran dentro de las veinticuatro horas si el contenido reportado por los usuarios como ilegal lo es realmente. El alcance del proyecto de ley se ha ampliado desde el 'discurso de odio' ilegal a una amplia gama de otros contenidos, que incluyen: apología a actos que constituyen un delito contra la dignidad humana, crímenes de guerra, crímenes de lesa humanidad,

esclavitud, delitos de colaboración con un enemigo, injerencia voluntaria en la vida o integridad física, agresión sexual, robo agravado, extorsión o destrucción, degradación o deterioro voluntario que sea peligroso para una persona, acoso sexual, trata de personas, proxenetismo, la incitación o la excusa para cometer actos de terrorismo y el contenido de abuso de menores. A diferencia de la ley alemana que sirvió de inspiración, la versión francesa no contó con dispositivos de análisis de contenido denunciados como ilegales durante un periodo prolongado en vista de la necesidad de una evaluación precisa.

El 18 de junio de 2020, el Tribunal Constitucional francés declaró inconstitucional gran parte de la ley (Conseil constitutionel, 2020). En su decisión afirmó que consideraba que determinadas disposiciones atentan contra "la libertad de expresión y comunicación y no son necesarias, adecuadas y proporcionales al fin perseguido". Se mantuvieron algunas de las normas legales previstas, menos de la mitad, incluido el permiso para crear un tribunal judicial especial dedicado al "odio en línea" y un observatorio del "odio en línea" diseñado para monitorear y analizar el contenido en línea y su desarrollo. El CSA actuará como secretaría de este observatorio.

2.3 CANADÁ - LEY DE MODERNIZACIÓN DE LAS ELECCIONES

En Canadá, la ley de modernización electoral (Bill C-76), aprobada en diciembre de 2018, cuenta con disposiciones que buscan dar respuesta a los desafíos que el entorno digital impone al proceso electoral. Sin embargo, la ley no trata en detalle ninguno de los aspectos de vanguardia: algoritmos de visibilidad, *microtargeting* de plataforma, cuentas no auténticas y comportamiento coordinado.

Las plataformas, que cumplen con los criterios de la ley, necesitan registrar anuncios, piezas de campaña o partidistas en su espacio publicitario (ya sea directo o indirecto), así como sus contratistas y agentes económicos (tanto de la plataforma como del contratista). Los criterios de consideración de las plataformas *online* varían según el idioma, como el número de accesos al mes: tres millones en inglés, un millón en francés y cien mil si el idioma no es uno de los anteriores.

Existen similitudes con medidas adoptadas en otros países, ya que la ley busca poner límites al gasto de los partidos y brindar más transparencia a la participación de terceros en el proceso electoral, que incluye actividades partidistas, campañas / propaganda y sondeos de opinión. O incluso en la protección de datos personales, ya que la propuesta obliga a las partes a publicar su política de protección de datos en sus sitios web (aplicada a los datos de los ciudadanos a los que las partes tienen acceso), bajo pena de perder su registro.

El reglamento se encarga de proteger a las personas que hablan de política en Internet sin fines comerciales, separándolas expresamente de los anunciantes en diferentes momentos. Sin embargo, incluso la distribución de enlaces se considera publicidad ("providing a link to an Internet page that does anything referred to in subparagraphs (i) and (ii)")[8], lo que abre la puerta a la falta de definición entre el intercambio orgánico y la publicidad.

2.4 INDIA

La India tiene en Information Technology Act 2000 - (IT Act), el Acto de tecnología de la información 2000, su principal referencia normativa para abordar el problema de la distribución de contenido falso y engañoso por las redes sociales y las plataformas de mensajería instantánea. El IT Act 2000 fue formulado, sin embargo, para regular el comercio y los delitos electrónicos y, en este contexto, trató como delito la publicación de informaciones tergiversadas y falsas (*misrepresentation*), de contenidos obscenos, de mensajes fraudulentos, así como la adulteración de documentos y la violación de privacidad, entre otras situaciones. Este estatuto faculta al gobierno a asumir el rol de juzgar qué es un delito cibernético y ha tenido artículos impugnados por restringir la libertad de expresión y la vigilancia masiva de la población. La ley se ha actualizado con enmiendas a lo largo del tiempo.

8 N. T.: "Proporcionar un enlace a una página de Internet que hace todo lo mencionado en los subpárrafos (i) y (ii)".

En 2015 se consideró inconstitucional el artículo 66A, que preveía sanciones, incluidas multas y penas de prisión, por enviar mensajes ofensivos por medio de los servicios de comunicación. Esto ocurrió después de que caricaturistas, políticos, empresarios y jóvenes habían sido arrestados y, o sus cuentas habían sido bloqueadas por publicaciones que critican al gobierno, especialmente en torno a la corrupción y los símbolos nacionales. A pesar de haber sido invalidada por la Corte Suprema, la Internet Freedom Foundation ha demostrado que los gobiernos estatales y la policía continúan usando el artículo 66A para bloquear cuentas y sitios que encuentran ofensivos[9].

El artículo 79, que en la ley vigente exime a los intermediarios de responsabilidad por los contenidos publicados por terceros, a su vez, está sujeto a modificaciones. En la más reciente, la Information Technology [Intermediary Guidelines (Amendment) Rules] 2018, incluye la obligación de los intermediarios de emitir condiciones de uso que prohíban a los usuarios publicar contenido "obsceno y dañino" y que amenace la "salud y seguridad públicas"; eliminar el acceso a contenido considerado ilegal, dentro de las veinticuatro horas, cuando el gobierno lo notifique, así como hacer uso de métodos automatizados para eliminar, por sí solo, contenido ilegal; brindar asistencia técnica al gobierno en un plazo de setenta y dos horas a partir de la solicitud y permitir el seguimiento del autor de la información en cuestión; establecer una sede en el país si tienen más de cinco mil usuarios o residentes indios en la India. Expertos, organizaciones y empresas tecnológicas consideran que modificaciones como estas atentan contra el derecho a la libertad de expresión (derecho fundamental previsto en el artículo 19 de la Constitución) y el derecho a la protección de los datos personales.

El borrador de esta enmienda fue presentado en diciembre de 2018 para consulta y manifestación de las partes interesadas. En 2020, una carta firmada por investigadores, activistas digitales y periodistas, entre otros, dirigida al Ministerio de Tecnología e Información Electrónica, conocido por las siglas MEITY, destacó como las principales preocupaciones "el uso desproporcionado de la regulación gubernamental", el carácter vago de términos que enfrentan la libertad de expresión, como "contenido extremadamente dañino, hostil y odioso"; y el problema de romper el cifrado que puede afectar la seguridad de los mensajes que circulan en aplicaciones como WhatsApp y la privacidad de los individuos[10]. La versión final aún no ha sido presentada por el gobierno y aprobada por el parlamento para convertirse en ley.

2.5 SINGAPUR

Singapur instituyó, en 2019, la ley de protección contra las falsedades y la manipulación en línea (Protection from Online Falsehoods and Manipulation Act), llamada también por la sigla Profma. Conocida como la "ley de las *fake news*" del país, el estatuto aprobado por el Parlamento en mayo (que entró en vigencia en septiembre de ese año) prohíbe la propagación en línea de hechos considerados falsos, al tiempo que crea medidas para neutralizar los efectos de estos mensajes. La ley prevé sanciones principalmente a las personas, pero también a los proveedores, en caso de incumplimiento de las decisiones provenientes del gobierno central y de los órganos de la administración directa, lo que es considerado abusivo por organizaciones como la Organización Internacional de Juristas, Reporteros sin Fronteras y Human Rights Watch.

Con el fin de evitar la circulación de falsedades y la manipulación de la información en línea, la ley denomina declaración falsa de hechos al acto de comunicar a uno o más usuarios finales, por Internet y, o SMS, declaraciones y materiales (incluidos mensajes, artículos, discursos, posts, imágenes, audios, videos, etcétera) que sean total o parcialmente falsos o engañosos. Con base en la ley, un individuo no puede publicar hechos y mensajes falsos que puedan ser perjudiciales para la imagen, la

9 Disponible en: <https://internetfreedom.in/how-a-bill-becomes-a-zombie-the-journey-of-section-66a-of-the-information-technology-act-2000/>. Consultado el 12 de noviembre de 2020.

10 Disponible en: <https://www.businesstoday.in/top-story/draft-information-technology-rules-experts-write-to-meity-highlight-key-concerns/ story / 317640.html>. Consultado el 12 de noviembre de 2020.

seguridad, la salud, las finanzas y la tranquilidad públicas, así como las relaciones amistosas y el clima electoral. Además, dentro del alcance del discurso del odio, no puede incitar sentimientos de enemistad, odio y mala voluntad entre diferentes grupos de personas, al igual que no puede disminuir la confianza pública en relación al gobierno.

Lo que se considera falso y engañoso, en el caso de asuntos como políticas públicas, elecciones, servicios públicos y órganos institucionales, depende de la valoración directa de la administración. El ministro, por ende, puede designar a la autoridad competente (consejo/dirección estatutaria, incluido el jefe de cualquier organismo al servicio del gobierno o de una autoridad estatutaria) para que desempeñe la función pública y ejecute las instrucciones del gobierno. Las personas pueden ser sancionadas con multas de hasta 100.000 dólares singapurenses (unos 75.000 dólares estadounidenses) y penas de prisión de hasta diez años.

La ley también tiene como objetivo extinguir cualquier financiación y promoción de declaraciones falsas de hecho con el apoyo de sitios en línea (*site*, página de internet, chat y foros, etcétera, alojados en una computadora y que se pueden ver, escuchar o percibir en Internet). Otro de los objetivos de este reglamento es el desarrollo de medidas para frenar cualquier comportamiento coordinado inauténtico y los usos indebidos de las cuentas en línea y los *bots*. Por ley, las actividades realizadas con dos o más cuentas para promover el engaño a las personas son conductas coordinadas no auténticas. Finalmente, la Profma prevé la creación de medidas para mejorar la difusión de información sobre contenidos pagados (en la ley, presentados como cualquier declaración comunicada en cualquier lugar con remuneración) con fines políticos.

Se pusieron en marcha tres medidas para garantizar el cumplimiento de la ley. La primera se llama Dirección de corrección (Correction Direction) y consiste en enviar un aviso de corrección a una persona, que debe seguir las pautas para informar públicamente que la declaración que él ha emitido es falsa o contiene alguna falsedad. "Una persona que haya comunicado una declaración de hecho falsa en Singapur puede recibir una instrucción de corrección, aunque no sepa o tenga razones para creer que la declaración es falsa", dice la letra de la ley[11]. La segunda se denominó Dirección para detener la comunicación (Stop Communication Direction), que determina que el individuo debe dejar de publicar declaraciones sobre un asunto determinado, o algo sustancialmente similar, en los términos especificados. Asimismo, la instrucción de interrumpir la comunicación puede ser válida incluso en los casos en que el individuo no sepa que esa declaración es falsa a la luz del entendimiento del gobierno.

La tercera medida se denomina Orden de bloqueo de acceso (Access blocking order), la cual es efectiva si el ciudadano no cumple con los avisos de rectificación e interrupción de la comunicación. En este caso, la ley habilita al ministro a ordenar que se deshabilite el acceso al lugar en línea donde efectivamente se comunica la declaración falsa a los usuarios finales. Mientras que en otros casos las penas como multa y encarcelamiento están dirigidas a la persona, en esta situación, el proveedor que no cumpla con la orden de bloqueo de acceso será declarado culpable y condenado al pago de una multa de veinte mil dólares singapurenses (unos US $15.000) por día, hasta un total de 500.000 dólares singapurenses (unos US $375.000).

Los afectados por las medidas pueden recurrir al Tribunal Superior, pero antes se le debe solicitar al ministro anular o cambiar las determinaciones. La ley establece que el Tribunal solo puede anular el acto cuando la persona no es, de hecho, responsable de comunicar la declaración; si no se establece una declaración de hecho o si la declaración de hecho es verdadera; si técnicamente no es posible cumplir la decisión. La orden permanece en vigor, incluso después de que se haya presentado una apelación, hasta que sea anulada por el Tribunal Superior o el Tribunal de Apelación (Court of Appeal), o si expira.

11 Traducción de: "A person who communicated a false statement of fact in Singapore may be issued a Correction Direction even if the person does not know or has no reason to believe that the statement is false". Disponible en: <https://docs.google.com/document/d/1W4aUzkQw_NFbr3TEwI-2HAWM6q-u8RmyA/edit>. Consultado el 12 de noviembre de 2020.

La primera orden de corrección se produjo en noviembre de 2019 y se dirigió al líder de la oposición Brad Bowyer (Partido del Progreso de Singapur) por una publicación en Facebook en la que cuestionaba la independencia de las empresas de inversión[12]. El aviso de que propagó información falsa y engañosa se publicó en la página de *fact-checking* (comprobación de hechos) del gobierno de Singapur[13]. Desde entonces, activistas y periodistas han sido advertidos y sancionados por publicar contenido y, o comentarios que contengan críticas a alguna acción y agencia gubernamental en sus propias páginas en plataformas de redes sociales. La ONG Human Rights Watch afirma en su Informe mundial 2020 que la libertad de expresión puede considerarse restringida en Singapur[14]. El contexto de Profma empujó al país de la posición 151 a la 158 en el Índice mundial de libertad de prensa de Reporteros sin Fronteras. Para la ONG,

> 2019 vio un deterioro significativo con la adopción de una ley de noticias falsas con disposiciones orwellianas que permiten al gobierno actuar como una combinación de Ministerio de la Verdad y oficina de censura en la era de las redes sociales[15].

2.6 UNIÓN EUROPEA - EU CODE OF PRACTICE ON DISINFORMATION (CÓDIGO DE PRÁCTICAS SOBRE DESINFORMACIÓN)

Uno de los mejores ejemplos de esfuerzos coordinados, que unen a empresas como Facebook, Twitter, Google, Microsoft y Mozilla, es el denominado Código de prácticas sobre desinformación de la UE (EU Code of Practice on Disinformation), de septiembre de 2018. Es un código de buenas prácticas para plataformas y asociaciones para prevenir o mitigar la difusión de información falsa, elaborado en respuesta a los problemas expuestos por la comunicación de la Comisión Europea en el documento "Tackling online desinformation: A European Approach" (Comisión Europea, 2018). El documento trata principalmente de la exposición de los ciudadanos europeos a información falsa o engañosa a escala masiva, donde la amplificación por parte de las redes sociales en línea (mediante algoritmos, anuncios dirigidos o robots) figura como una parte clave en las causas del problema. Las soluciones propuestas por la Comisión incluyen más transparencia, diversidad de información, indicios de credibilidad (etiquetas de contenido verificado o que está en disputa) y alfabetización mediática.

El código de buenas prácticas suscrito por las empresas busca también la compatibilidad con otras propuestas y normas, como la Carta de los derechos fundamentales de la Unión Europea, la Convención europea de derechos humanos, la normativa de la UE 2016/679 sobre la protección de las personas en lo que respecta al tratamiento y movimiento de sus datos personales, y las directivas 2000/31 / CE (que se ocupa del comercio electrónico, en particular los artículos 12 a 15, que tratan de las garantías que deben dar los Estados miembros, asegurando que la prestación de servicios preserva la confidencialidad de las comunicaciones y evita el almacenamiento o el uso indebido), 2005/29 / EC (sobre prácticas comerciales desleales entre empresas y consumidores) y 2006/114 / EC (que trata, entre otras cosas, sobre publicidad engañosa).

El código establece conductas y buenas prácticas para hacer frente a la difusión de información falsa en línea, comprometiéndose a (a) ser más cuidadoso con la distribución de anuncios dirigidos que incluyan información falsa, b) hacer más transparentes los contratistas y los importes pagados en los anuncios que impliquen asuntos políticos o controversias específicas (publicidad basada en temas), (c) la integridad de los servicios evitando las cuentas falsas o el uso inadecuado de robots y, por último, empoderar (d) a los consumidores y (e)

12 Disponible en: <https://docs.google.com/document/d/1FbUvVFWW2mE1p6vR8ATVqNGRT2G_Cdfoq-JBW7Whoas/edit#>. Consultado el 20 de noviembre de 2020.

13 Disponible en: <https://www.scmp.com/week-asia/politics/article/3039260/singapore-invokes-fake-news-law-first-time-over-politicians>. Consultado el 20 de noviembre de 2020.

14 Disponible en: <https://www.hrw.org/world-report/2020/country-chapters/singapore>. Consultado el 20 de noviembre de 2020.

15 Traducción de: "2019 saw a significant deterioration with the adoption of an anti-fake news law with Orwellian provisions that allows the government to act as a combination of Ministry of Truth and censorship office for the social media era". Disponible en: <https://rsf.org/en/singapore>. Consultado el 20 de noviembre de 2020.

a la comunidad académica/investigadora. Los firmantes facilitarían informes anuales de evaluación y acciones que los firmantes pueden poner en marcha para hacer frente a la desinformación. El documento suscrito por las empresas es un código esencialmente proposicional y de principios, sin claras imposiciones ni sanciones, y los firmantes pueden dejar los compromisos sin mayores consecuencias.

En el primer informe anual (octubre de 2019) se abordaron las políticas sobre anuncios y número de anuncios que dieron lugar a respuestas de la plataforma, exposición de los compradores de los anuncios mostrados a los usuarios, eliminación y bloqueo de cuentas con comportamiento perverso, notificación a los usuarios y alianzas con verificadores de datos (*fact-checkers*) para valorar el contenido verificado y las iniciativas que implican facilitar el acceso a datos para los investigadores.

2.7 REINO UNIDO

El gobierno del Reino Unido ha estado trabajando y buscando soluciones al problema relacionado con el contenido ilegal y nocivo en línea. En abril de 2019, la Secretaría de Estado de Cultura Digital, Medios de Comunicación y Deporte y la Secretaría de Estado del Ministerio del Interior presentaron un análisis del contexto, el impacto y los primeros pasos para forjar medidas reguladoras. Después de ser presentado públicamente, el *Libro blanco de daños en línea* (*Online Harm White Paper*) se puso a consulta pública. En febrero de 2020 se dio a conocer el informe con la sistematización de aportes.

Este informe contiene la intención del ejecutivo de enviar un proyecto de ley que cree un "deber de cuidado" para las empresas que operan en servicios que permiten la interacción con contenidos generados por terceros o la interacción entre usuarios.

Este deber de cuidado se detallará en los códigos que se elaboren, señalando cada uno de los temas que apruebe el Ministerio del Interior, y será aplicado y supervisado por la Oficina de Comunicaciones (Ofcom), el regulador de los servicios de comunicaciones, que en la propuesta obtendrá nuevos poderes. Las empresas podrán presentar formas alternativas a las definidas en los códigos y deberán justificar por qué serían más eficientes y adecuadas. Las prioridades para la producción de códigos son proteger a los niños y adolescentes de la explotación sexual y combatir los contenidos que representan una amenaza para la seguridad nacional y el terrorismo. En cuanto al tratamiento del fenómeno de la desinformación, el documento prevé que las empresas tomen medidas proactivas y proporcionadas para ayudar a los usuarios a comprender la naturaleza y confiabilidad de la información que están recibiendo, pero sin mayores detalles.

Cabe mencionar que el gobierno señaló, en la consulta pública, y de manera genérica, la intención de regular los servicios privados de comunicación, a lo que se opusieron los encuestados. Sin embargo, dice el gobierno, se ha reconocido en algunas respuestas –tanto de individuos como de organizaciones–, que el abuso, el acoso y algunas de las actividades ilegales más graves ocurren en espacios privados, como foros comunitarios cerrados y salas de chat. La propuesta presentada indica que Ofcom tendrá una serie de poderes para tomar medidas efectivas contra las empresas que incumplan el deber legal de atención. Esto puede incluir el poder de emitir multas sustanciales e imponer responsabilidades a las personas que están a cargo de la alta dirección corporativa.

El regulador estará facultado para exigir informes anuales de transparencia a las empresas incluidas en el ámbito de aplicación, en los que se expondrá la prevalencia de los contenidos en sus plataformas y las medidas adoptadas. Estos informes se publicarán en línea para que los usuarios y los padres puedan tomar decisiones informadas sobre el uso de Internet. El regulador también estará facultado para exigir información adicional, incluso sobre el impacto de los algoritmos en la selección de contenido para los usuarios, y para garantizar que las empresas informen proactivamente sobre daños emergentes y conocidos.

El regulador fomentará y supervisará el cumplimiento de los compromisos de las empresas con los investigadores independientes, así como las garantías adecuadas para

proteger a los usuarios. También supervisará la disponibilidad de canales para los reclamos de los usuarios y la respuesta a las quejas. Se está evaluando el permiso de algunos organismos especiales para hacer "superreclamaciones" al regulador.

Los grupos de derechos humanos han expresado su preocupación de que el enfoque de aplicación propuesto pueda ser desproporcionadamente punitivo, y el regulador tendría que demostrar que ha cumplido con la prueba de proporcionalidad para tomar medidas para frenar la libertad de expresión bajo las leyes de derechos humanos. La preocupación es mayor debido a la previsión de bloqueo de los servicios *online* directos por parte de los proveedores de conexión. Además, existe un temor generalizado de que la presión en torno a las empresas por soluciones pueda conducir a una restricción privada de la libertad de expresión, como precaución contra un posible castigo. En el caso de medidas para contener fenómenos cuyo concepto es bastante abierto, como la desinformación, existe el temor de restringir el ejercicio legítimo de la libertad de expresión.

El *Libro blanco* (*White paper*) se publicó en 2019 y el informe de consulta a principios de 2020[16].

2.8 BRASIL - PROYECTO DE LEY 2630/2020

En Brasil, un proyecto de ley aprobado en el Senado, el PL 2630/2020, del senador Alessandro Vieira, pretende regular las redes sociales y los servicios de mensajería instantánea para contener el fenómeno de la desinformación y garantizar una mayor transparencia sobre cómo las plataformas han venido gestionando los contenidos. El texto fue aprobado en el Senado Federal en junio de 2020, pero al momento de cerrar este texto [diciembre de 2020] aún no había sido revisado por la Cámara de Diputados.

La propuesta prevé la creación de una entidad de control para acompañar la aplicación del reglamento, encargada también de detallarlo en un código de conducta, con participación multisectorial, y establecida en el seno de la legislatura. El PL también incluye una serie de obligaciones en materia de promoción y publicidad, especialmente electoral, y prohibición de robots no identificados, además de un informe de transparencia con una serie de requisitos. Existen disposiciones para el almacenamiento de la identificación de los anunciantes y el almacenamiento de metadatos de los mensajes privados que se comparten en grupos y que alcanzan un determinado nivel de compartición. Además, prevé una enmienda a la ley que rige el servicio móvil privado (teléfono celular) para exigir a las empresas de telecomunicaciones que validen los datos registrados por los usuarios.

Ante la realidad nacional, en la que los miembros del Poder Ejecutivo y Legislativo son considerados importantes difusores de desinformación y agresiones, la propuesta se centra en el uso de las redes sociales por parte de los poderes públicos y de los servidores públicos.

La redacción votada por el Senado incorporó puntos sobre el debido proceso en la moderación de contenidos por parte de las plataformas, como los mecanismos de notificación y el derecho de defensa de los usuarios, que son importantes para el ejercicio de la libertad de expresión. En caso de que la aplicación de la moderación se considere inadecuada, corresponderá a la plataforma reparar los daños resultantes de la interferencia.

El proyecto también prevé acciones rápidas por parte de las plataformas y reacciones a las quejas, ofreciendo el derecho de respuesta, en un formato vago y poco claro. El texto establece que la decisión del procedimiento de moderación debe garantizar "a la víctima el derecho de réplica en la misma medida y alcance de los contenidos considerados inapropiados".

Uno de los puntos más debatidos y que divide las opiniones es el almacenamiento *a priori* de los metadatos de los mensajes en los servicios de mensajería. La propuesta es mantener los metadatos de aquellos mensajes que alcancen un nivel de compartición y en grupos, con el fin de posibilitar, por orden judicial, la identificación de los responsables de los mensajes considerados ilegales. La mayoría de las entidades de la sociedad civil abogan por

16 Nota del editor: en diciembre de 2020, luego del cierre de este documento, se publicó una nueva versión aún más detallada de la respuesta del gobierno a la consulta.

la supresión total del artículo 10, y WhatsApp, una aplicación de Facebook, presentó la opción de almacenar metadatos de interacciones solo después del inicio de una investigación. Otras organizaciones de la sociedad civil defienden el artículo partiendo de la lectura de que la imposibilidad actual, en la práctica, de la rendición de cuentas legal de los responsables de contenidos ilícitos, funciona como un incentivo para la práctica de la desinformación.

3

ANÁLISIS TRANSVERSAL DE LOS CASOS

La valoración de los textos legales presentados en el capítulo anterior permite identificar cuestiones transversales que ofrecen una clave de lectura sobre cuáles son los puntos críticos para el abordaje regulador de la desinformación.

Esta sección proporciona una descripción general de los principales puntos identificados. La tabla 1 (página 20) muestra cómo responde cada caso a cada uno de estos puntos críticos.

1) *Definición y arbitrio de veracidad*. La definición de veracidad, o el criterio que separa la información falsa de la verdadera, es un nodo central en los debates sobre desinformación. Hasta el día de hoy, en un escenario de predominio del periodismo profesional, estas definiciones, por regla general, las daban sobre todo los redactores de los medios y, en caso de disputa, el Poder Judicial. Hasta hace poco, las plataformas no asumían el lugar de los editores de contenido y no arbitraban sobre la veracidad del contenido publicado por terceros. Sin embargo, el volumen y el impacto del fraude de información y de contenidos ilícitos comenzaron a suscitar el debate sobre la necesidad de que las plataformas, o agencias de verificación, tengan el poder de arbitrar sobre la veracidad de la información. Este modelo ha sido adoptado en los últimos años y justifica, en varios casos, la eliminación de contenido o la reducción de alcance. La opción de trabajar con el arbitraje privado sobre la verdad genera inquietudes en relación a la libertad de expresión de los usuarios, ya que se empieza a atribuir el rol de juez a un actor privado. Además de estos dos escenarios, también existen otros casos más críticos en los que el gobierno asume el rol de árbitro, como en Singapur, en el que la ley otorgó al órgano de administración directo, vinculado al Poder Ejecutivo, el rol de definir la veracidad de cierta información.

En los textos analizados, aunque las *fake news* aparecen como expresión de fachada y, aunque la información, el contenido y las falsas declaraciones aparecen como uno de los principales objetivos de las normas legales, no siempre se presenta una definición. Al mismo tiempo, algunas de estas leyes y códigos no regulan necesariamente las comunicaciones falsas, sino que eligen enmarcar el discurso de odio específicamente como un delito para evitar que los extremistas de todo tipo prosperen en línea.

2) *Responsabilidad de la aplicación.* Los modelos analizados combinan, en distintas proporciones, sistemas de autorregulación, corregulación y regulación pública. En la autorregulación, la definición de criterios y su aplicación son responsabilidad de las empresas. En la corregulación, la definición de criterios se da por ley, la aplicación inicial la hacen las empresas y los organismos públicos supervisan la aplicación. En la regulación pública, los criterios vienen dados por ley o por regulación infralegal y la solicitud la hace directamente un organismo público. Cada uno de estos modelos tiene ventajas y desventajas. En cualquier caso, la ausencia de criterios públicos de moderación de contenidos repercute en derechos fundamentales de los usuarios tanto cuando la solicitud la hace la propia empresa como cuando la hace un organismo regulador.

3) *Atribución y tipo de responsabilidad legal*. Algunos textos legales asignan responsabilidad legal solo a los usuarios responsables de contenidos falsos o engañosos, otros también asignan responsabilidad a las plataformas. En el caso de la responsabilidad de los usuarios,

esta puede ser civil o penal. En el caso de la responsabilidad de la plataforma, esta se puede dar en tres niveles: primero, en cada uno de los casos individuales en los que exista una vulneración de derechos o práctica ilícita por parte de los usuarios y no haya habido actuación de la plataforma. En un segundo nivel, sobre un deber general de cuidado, como proponen Alemania y el Reino Unido, sin responsabilidad para todos y cada uno de los casos. O, inclusive en un tercer nivel, con responsabilidad solo en los casos en que la plataforma incumpla una orden judicial, un modelo que prevalece hoy en Estados Unidos, la Unión Europea y Brasil. Este tema se detallará en la siguiente sección.

4) *Enfoque preferencial*. Los textos legales aportan diferentes enfoques al problema en un intento de enfrentar la desinformación. En algunos casos, la atención se centra en el contenido falso o engañoso. Un segundo enfoque intenta incidir sobre el comportamiento de los usuarios, buscando prevenir, por ejemplo, los robots no identificados en Twitter, la omisión de la identidad del responsable de una determinada página de Facebook o la distribución de correos masivos en WhatsApp. Un tercer enfoque es buscar incidir en el financiamiento de la desinformación, por ejemplo, mediante la transparencia sobre la responsabilidad de los anuncios políticos. Un cuarto enfoque posible es cuando se busca afectar la arquitectura de las redes sociales, por ejemplo, imponiendo a las empresas la obligación de almacenar datos.

5) *Alcance y jurisdicción*. Las características que definen a Internet como de alcance global, su carácter privado y la propensión a la innovación y la digitalización todavía se materializan ante los reguladores como dificultades en la adhesión de los mecanismos legales tradicionales a las interacciones virtuales.

Es en este escenario que varias normativas inician experiencias de regulación de contenidos en plataformas que utilizan contenidos generados por terceros. Cabe destacar, sin duda, el intento del Estado nacional de hacer cumplir su ley, siendo la obligación de designar un responsable legal para la empresa una exigencia que se repite en varios ordenamientos legales y propuestas normativas aquí analizadas. La estrategia tiene como objetivo abordar el hecho de que Internet permite la prestación de un servicio global sin la necesidad de una presencia formal en las localidades.

Otra estrategia para enfrentar este desafío que plantea el modelo descentralizado de prestación de servicios en línea es la regulación con extraterritorialidad, afirmando el poder del Estado para ir más allá de sus territorios en determinadas condiciones. En Singapur, la ley para combatir las declaraciones consideradas falsas permite a un tribunal citar a una persona que ha cometido un delito fuera del país, como si el delito se hubiera perpetrado en el territorio.

La Internet Society, un documento que analiza pronósticos legales con impacto extraterritorial, sugiere que los gobiernos busquen dialogar con otras (*stakeholders*) *partes interesadas* para obtener el resultado esperado (Internet Society, 2018). De hecho, iniciativas multisectoriales y transnacionales podrían servir de referencia para acciones más efectivas. Un buen ejemplo es la iniciativa Christchurch Call[17], encabezada por los gobiernos de Francia y Nueva Zelanda, con el apoyo inicial de otros quince países[18], más la Comisión Europea. La iniciativa fue creada en respuesta a los ataques terroristas del 15 de marzo de 2019 en la comunidad musulmana de Christchurch, en Nueva Zelanda. Christchurch Call consiste en una serie de propuestas de compromisos voluntarios para que los gobiernos y las redes sociales actúen de manera coordinada y sistemática para contener contenido extremista violento en línea y prevenir el abuso de Internet por parte de organizaciones e individuos terroristas. Aunque se basa en compromisos genéricos, la convocatoria ha logrado una adherencia significati-

17 <https://www.christchurchcall.com/call.html>

18 Fundadores, Nueva Zelanda y Francia. Partidarios fundadores: Australia, Canadá, Comisión Europea, Francia, Alemania, Indonesia, India, Irlanda, Italia, Japón, Jordania, Holanda, Nueva Zelanda, Noruega, Senegal, España, Suecia y Reino Unido. Apoyadores: Argentina, Austria, Bélgica, Bulgaria, Chile, Chipre, Colombia, Corea del Sur, Costa de Marfil, Costa Rica, Dinamarca, Eslovenia, Finlandia, Georgia, Ghana, Grecia, Hungría, Islandia, Kenia, Letonia, Lituania, Luxemburgo, Maldivas, Malta, México, Mongolia, Polonia, Portugal, Rumania, Sri Lanka, Suiza, Unesco, Consejo de Europa. Proveedores de servicios: Amazon, DailyMotion, Facebook, Google, Microsoft, Qwant, Twitter, YouTube, Line.

va. En septiembre de 2019 se unieron otros treinta y un países, además de la Unesco, así como los principales proveedores de servicios en línea como Amazon, Google, Facebook, YouTube, Twitter, Microsoft, Qwant, Line, DailyMotion y JeuxVideos. Una iniciativa similar podría llevarse a cabo sobre la desinformación.

6) *Grado de seguimiento de la acción de las plataformas por parte del gobierno o de investigadores independientes*. Diferentes textos legales establecen obligaciones de transparencia para las empresas, que sirven no solo para permitir al usuario conocer y monitorear las políticas de moderación de contenidos sino también para evaluar la eficacia de la aplicación de las leyes y normas reguladoras. En el caso de Alemania, por ejemplo, Facebook fue multado porque, entre otras cosas, el número de denuncias registradas de infracciones contra NetzDG era mucho menor que en otras plataformas. En Brasil, durante el proceso electoral de 2020, Facebook afirmó haber eliminado más de 140.000 contenidos por violar las políticas de interferencia electoral. Sin acceso a los datos por parte de investigadores independientes es difícil saber si las empresas están utilizando criterios legalmente definidos, si están interfiriendo en los contenidos legítimos o si están siendo negligentes.

Todos estos puntos críticos contienen tensiones entre derechos fundamentales, siendo los más comunes: la libertad de expresión individual frente al acceso a la información (entendida como información plural, diversa y fiable), y la libertad de expresión individual frente a la privacidad y la protección de datos. Al abordar la regulación para contener el fenómeno de la desinformación, es necesario tener en cuenta la obligación de los gobiernos de garantizar el pleno goce de la libertad de expresión y acceso a la información, ya que estos derechos están directamente relacionados.

La Corte Interamericana de Derechos Humanos (CIDH) ha expresado que:

> la libertad de expresión es la piedra angular de la existencia misma de una sociedad democrática. Es fundamental para la formación de la opinión pública. También es una condición *sine qua non* para que los partidos políticos, los sindicatos, las sociedades científicas y culturales y, en general, quienes deseen influir en el colectivo puedan desarrollarse plenamente. Es, en definitiva, una condición para que la comunidad, a la hora de ejercer sus opciones, esté suficientemente informada (Corte IDH, 1985).

Así, es posible afirmar que una sociedad que no está bien informada no tiene total libertad para expresarse. En este sentido, la búsqueda de un ciclo virtuoso entre la libertad de expresión y el derecho a estar bien informado debe ser perseguida por la política pública.

Comprender la necesidad de regulación no significa que se comprenda inmediatamente el esquema adecuado para lograr el objetivo deseado. En el caso de la regulación de Internet y el entorno digital, cabe recordar la teoría desarrollada por Lawrence Lessig (2006) de que, además de la ley, el mercado y las normas sociales, el propio código de programación (la arquitectura) tiene fuertes efectos reguladores.

Tabla 1

	Ley NetzDG (Alemania)	Ley 2018-1202 (Francia)	Ley de modernización de las elecciones (Canadá)	Ley de tecnología de la información (India)	Ley de protección contra la falsedad y la manipulación (Singapur)	Código de prácticas sobre desinformación (Unión Europea)	Libro blanco sobre daños en línea (Reino Unido)	Proyecto de ley para luchar contra las *fake news* 2630/20 (Brasil)
Aprobación	**2017**	**2018**	**2018**	**2000/2018**	**2019**	**2018**	**En discusión**	**En discusión**
Definición y arbitraje sobre la veracidad	Exige la acción directa de las plataformas contra contenidos 'evidentemente ilegales' (delitos ya previstos en el Código penal alemán).	Faculta (y cobra) a las plataformas para actuar y arbitrar directamente.	No. La atención se centra en la transparencia de la financiación publicitaria en el proceso electoral.	Sí, por administración directa.	Sí, por administración directa.	Faculta a las plataformas para actuar y arbitrar directamente.	Faculta (y cobra) a las plataformas para actuar y arbitrar directamente.	No crea reglas al respecto.
Responsabilidad de la aplicación (gubernamental, autorregulación o corregulación)	Co-regulación (autorregulación regulada).	Corregulación (control por parte del Consejo Superior Audiovisual), con un régimen especial para el proceso electoral.	Autoridad electoral.	Regulación gubernamental.	Regulación gubernamental.	Autorregulación.	Corregulación (criterios en la ley, aplicación por parte de las plataformas y supervisión por parte de Ofcom).	Corregulación (reglas definidas por ley aplicadas por las plataformas), con supervisión del Consejo vinculado al Congreso.
Atribución y tipo de responsabilidad legal (personas físicas o plataformas)	Impone una especie de "deber de cuidado" a las plataformas.	Impone una especie de "deber de cuidado" a las plataformas.	Impone obligaciones a campañas, partidos, candidatos y simpatizantes.	Exime a las plataformas de responsabilidad por el contenido de terceros, pero la enmienda propuesta impone la responsabilidad de eliminación y bloqueo.	Proporciona duras sanciones a personas y plataformas.	No hay. Crea un código de prácticas para la adhesión voluntaria a las plataformas, sin sanciones.	Impone una especie de "deber de cuidado" a las plataformas. El foco está en la responsabilidad de las plataformas.	Impone a las plataformas obligaciones equivalentes a un deber de cuidado. No impone sanciones a los usuarios.
Enfoque preferencial	Contenido.	Comportamiento de los usuarios y financiación.	Financiación.	Contenido y arquitectura de las plataformas.	Comportamiento y contenido de los usuarios.	Comportamiento y contenido de los usuarios.	Contenido y arquitectura de las plataformas.	Comportamiento de los usuarios y arquitectura de las plataformas.
Alcance y jurisdicción	Nacional, con impacto extraterritorial.	Nacional.	Nacional.	Nacional.	Nacional, con impacto extraterritorial.	Regional (Unión Europea).	Nacional.	Nacional.
Transparencia / seguimiento por parte de investigadores independientes	Impone importantes obligaciones de transparencia en las plataformas.	Sí, en lo que respecta a los impulsos, el valor y los datos utilizados. Además, propone transparencia de los algoritmos.	Centrarse en la transparencia del gasto en anuncios electorales en las plataformas.	No.	No.	Sí, sobre anuncios políticos y algunas medidas generales para plataformas.	Sí, el órgano regulador puede solicitar un informe anual, información y explicación del impacto de los algoritmos.	Sí, impone la obligación de transparencia de las plataformas, incluidos el impulso y la publicidad.

4

RESPONSABILIDAD DE INTERMEDIARIOS Y USUARIOS

El asunto que históricamente ha sido central en la organización del debate regulador sobre los contenidos en línea es la responsabilidad de los intermediarios. La aprobación, en 1996, en Estados Unidos, de la "Communications Decency Act" (DCA), que surgió luego del intento de regular la pornografía *online*, estableció un modelo de regulación para las plataformas digitales, considerándose aquí los negocios que conectan una o varias partes de dichas transacciones, utilizando contenidos producidos por terceros.

La sección 230 de la DCA pasó a proteger los sitios web y posteriormente las plataformas: "Ningún proveedor o usuario de un servicio de computación interactiva debe ser tratado como el editor o hablante de cualquier información proporcionada por otro proveedor de contenido de información". La sección no solo protege a los sitios web y plataformas de ser legalmente responsables por el contenido de terceros, sino que permite "cualquier acción tomada voluntariamente de buena fe para restringir el acceso o la disponibilidad de material que el proveedor o usuario considere obsceno, lujurioso, sucio, excesivamente violento, hostil o de alguna forma cuestionable", sin ninguna responsabilidad por sus decisiones. Esta amplia autorización (especialmente marcada por la expresión 'de alguna forma cuestionable') ofrece la posibilidad de aplicar sus reglas comunitarias sin presentar ninguna justificación para las expulsiones y sin condiciones.

El extracto de la ley, que ha llegado a ser referida como "la herramienta más importante para la libertad de expresión en Internet" o "la ley más importante en Internet", creó un estándar, que fue adoptado de manera adaptada, con algunas diferencias, en la directiva de comercio electrónico de la Unión Europea, en 2000, y en el Marco civil de Internet de Brasil, en 2014, entre otros países. Los usuarios y los servicios en línea utilizados no se confunden, por lo que se creó una seguridad jurídica para que los contenidos publicados no tuvieran que ser analizados constantemente por las entonces incipientes empresas y, ante cualquier amenaza de riesgo legal, fueran retirados del aire. Esta protección permitió que las plataformas crecieran sin ser molestadas en todo momento por personas insatisfechas con el contenido publicado por sus usuarios y sin generar un efecto silenciador sobre la libertad de expresión de los usuarios.

Hay que tener en cuenta que la opción del legislador estadounidense no pudo crear un entorno de total libertad de expresión sin considerar la necesidad de contener ciertos discursos. Por el contrario, al evitar la responsabilidad objetiva sobre los contenidos generados por terceros, preveía permitir a los intermediarios hacer todo lo posible para evitar contenidos nocivos e ilegales y no tener que elegir entre: 1) intentar moderar el contenido de terceros y hacerse responsable objetivamente cuando fallara, o 2) renunciar por completo al intento de frenar el abuso.

Sin embargo, la elección del régimen de responsabilidad civil elegido es señalada como el pilar de la baja inversión y la lenta reacción de las plataformas ante las situaciones externas negativas de sus negocios, incluido el fenómeno de la desinformación a escala global, explorada especialmente por grupos políticos, que utilizan información falsa o engañosa para disputar las posiciones de los ciudadanos.

En este sentido, se observa globalmente un intento de encontrar un nuevo régimen de responsabilidad para las plataformas digitales respecto a los contenidos produci-

dos por terceros. En el caso de las propuestas de Reino Unido y Brasil y de las leyes de Canadá y Alemania, se puede afirmar que existe un esfuerzo por señalar lo que se espera de estas empresas en cuanto a hacer el mejor esfuerzo para contener contenidos nocivos, sin que sean responsabilizadas de cada contenido producido por sus usuarios. Cabe señalar que algunas de las regulaciones identificadas como destructoras del estándar de la Sección 230 de la CDA estadounidense no lo son. Lo que se crea es una responsabilidad sistemática para tomar medidas preventivas y reaccionar rápidamente ante los contenidos ilegales o perjudiciales. Esto no significa que, ante la inacción en cuanto a un contenido específico, se produzca un cambio en el régimen de responsabilidad civil.

Algunos gobiernos, sin embargo, buscan soluciones para promover una mayor proactividad de las plataformas para evitar que sus actividades económicas afecten a los derechos garantizados por la ley y permitan la perpetración de crímenes. En términos generales, este segundo modelo puede ejemplificarse con las regulaciones de India y Singapur. Las mencionadas leyes, al promover la gestión individual de contenido a partir de demandas gubernamentales basadas en conceptos genéricos de desinformación –sin necesidad de la participación de la justicia u organismo independiente– no observan los principios de necesidad, proporcionalidad y legalidad que orientan cualquier iniciativa legítima para restringir la libertad de expresión por estándares internacionales, y pueden ser calificadas como normas para la aplicación de la censura estatal.

Mientras que la protección de las plataformas frente a la responsabilidad por los contenidos de terceros es importante para evitar el efecto silenciador, la cuestión de la responsabilidad de los individuos por los contenidos difundidos en la red está lejos de haber encontrado un punto de equilibrio. Varios análisis críticos acerca de la respuesta a contenidos de odio, por ejemplo, señalan la baja capacidad de los estados y los sistemas de justicia para castigar a los autores de delitos relacionados con calumnias, injurias, difamación, racismo, homofobia, transfobia, amenazas, etcétera. En este sentido, sigue siendo pertinente la búsqueda de modelos que puedan proteger la libertad de expresión y, al mismo tiempo, ofrecer mecanismos eficaces para defender los demás derechos humanos con los que pueda entrar en colisión. Quizás el mayor desafío sea cómo garantizar que las respuestas contra los contenidos ilegítimos sean adecuadas a su enorme volumen y a su rápida velocidad de difusión.

5

CUESTIONES AUSENTES

Mientras que la sección anterior analizó las cuestiones transversales presentes en los diferentes textos legales, esta sección propone discutir las faltantes. Cuestiones con potencial para incidir en la práctica de la desinformación, pero que han sido descuidadas o dejadas de lado en las definiciones normativas.

5.1 PROTECCIÓN DE DATOS

Si nos fijamos solo en el contenido de la manipulación informativa, lo que da sentido a estas estrategias queda en un segundo plano: la orientación de esta información a los grupos más proclives a compartirla, ayudando a la viralización y apropiándose de los criterios de visibilidad segmentada de los algoritmos en las plataformas. La misma información, inicialmente divulgada en grupos con diferentes perfiles, puede desencadenar procesos virales o ser completamente ignorada (Margetts et al., 2016). Es necesario tener en cuenta el hecho de que los contenidos no se mueven de forma autónoma y la pluralidad de formas en las que las personas pueden reaccionar al contacto con ellos.

Considerando el flujo de esta información y la precisión de sus objetivos, la protección de datos personales es una de las pocas acciones capaces de interferir, al mismo tiempo, en el procesamiento de datos por parte de agencias que venden servicios de segmentación irregular para mensajes criminales en WhatsApp, en la expansión de la visibilidad para nichos específicos siguiendo el funcionamiento de los algoritmos de plataformas como Facebook y en el uso de estos datos para la venta de publicidad dirigida en Instagram u otras redes. Esto se debe a que la protección reconoce que los ciudadanos son titulares de sus propios datos (que en este caso no pueden ser tratados para fines para los que no han prestado su consentimiento) y exige medidas de protección y transparencia en los casos en que los tratamientos estén autorizados.

Llegados a este punto, cabe recordar la comparación realizada por Bimber et al. (2012): de la misma forma que el impacto de los coches en el urbanismo provoca que todos aquellos que no poseen coche se vean afectados en la distribución de las calles, la limitación de aceras, la velocidad de los flujos de la ciudad, las políticas de transporte, etcétera, el impacto social de internet supera el simple acceso a la red. Lo mismo ocurre con los datos personales. Además de los productos de comunicación, los dispositivos y plataformas que utilizan datos personales permean otras dimensiones cotidianas –operaciones bancarias, relaciones profesionales, alimentación, transporte– involucradas también en la producción descentralizada de datos que son de gran utilidad para la identificación de perfiles y la comunicación política direccionada.

5.2 ARQUITECTURA DE LAS PLATAFORMAS Y APLICACIONES

El modelo de negocio de las aplicaciones se basa en la perspectiva de acumular millones de puntos de datos sobre cada usuario para permitir la creación de perfiles de usuario, con el fin de ofrecer perfiles hipersegmentados a los anunciantes. Así, el modelo de negocio basado en la acumulación de datos genera fragmentación y los algoritmos y la inteligencia artificial se guían por valores más relacionados con la uniformidad y la similitud que con el pluralismo y la diversidad. Esta contradicción está directamente relacionada con el objetivo de mantener al

usuario en la plataforma por más tiempo. Se deriva de la arquitectura y las disposiciones de la infraestructura de las redes sociales, que repercuten en el flujo y la jerarquización de los contenidos informativos, las ideas y las opiniones. Así, la fragmentación y la segmentación son fenómenos que se refuerzan en espiral y refuerzan lo que los estudiosos han denominado filtros de burbujas y cámaras de eco (Pariser, 2011).

Otro punto que está en segundo plano son las diferencias drásticas entre la economía política de los actores involucrados en las plataformas y redes sociales y la economía política en la comunicación audiovisual. Las plataformas unifican en una sola red social en línea los canales relacionados con las distintas industrias de la comunicación audiovisual –radiodifusión, cine, música, prensa y publicaciones electrónicas, juegos electrónicos o digitales, *marketing*, relaciones públicas, publicidad–, mientras que los contenidos disponibles en línea se producen de forma descentralizada y se distribuyen de forma segmentada (*narrowcast*). Diferentes industrias entran en un nicho de competencia, disputando las recomendaciones de los algoritmos de visibilidad y la atención de los usuarios, en un choque en el que los "*influencers*" individuales aportan nuevos elementos al debate.

Por un lado, los algoritmos cambian la propensión a indicar videos en función de la afinidad de los usuarios que han tenido contacto con estos videos; por otro, las apps pueden ser utilizadas en campañas que canalizan los accesos y la participación en otras plataformas, trazando interconexiones complejas entre los usuarios, las diferentes plataformas y sus algoritmos. Varias plataformas conforman un mismo sistema de redes interconectadas, donde los cambios en un punto pueden generar procesos de adaptación en otras redes y no se puede cambiar un punto esperando que el resto permanezca estático. La concentración de la propiedad se refiere, por lo tanto, a la concentración del poder para tomar decisiones como los criterios de los algoritmos, las políticas de uso y la privacidad de los datos (puntos marginales en los debates sobre los medios de difusión), y no a la concentración de los actores capaces de difundir contenidos en un canal específico. También concentra el poder a la hora de definir los límites y la responsabilidad de los profesionales que escriben algoritmos que tomarán decisiones (qué contenido priorizar, qué contenido debe mostrarse lo menos posible, qué contenido es una amenaza potencial), reproduciendo y acentuando a menudo las desigualdades estructurales.

5.3 OPACIDAD Y VIRALIZACIÓN

Por otro lado, las aplicaciones de mensajería sin algoritmos de visibilidad tienen un papel central en las campañas de desinformación. Destacamos cuatro aspectos: 1) en ausencia de una línea de tiempo, las publicaciones se almacenan en los dispositivos de los usuarios (mientras que eliminar una publicación con un millón de compartidos en Facebook elimina la fuente de esas acciones, en los mensajes de WhatsApp con un millón de reenvíos, para no hablar de los grupos, están en un millón de dispositivos diferentes, que pueden volver a insertarlos en otras redes en cualquier momento; 2) garantizan que la información estará expuesta a todos los miembros de los grupos a los que llega esta información (sin importar cuán desinteresados están en relación con la política o lo mal que han reaccionado en contactos anteriores), repetidamente; 3) la combinación entre reenvíos y grupos permite un aumento exponencial de la visibilidad propensa a la viralización, especialmente considerando la posibilidad de construir grupos segmentados a partir de datos personales tomados de otras redes (si cada miembro de un grupo de 256 personas lo reenvía a un nuevo grupo completo, se llegará a 65.500 personas en la primera ronda de reenvíos y a 16,7 millones en la segunda); 4) finalmente, el patrón de mensajes privados (encriptados y opacos por razones de seguridad) hace que se limite la investigación y seguimiento de estas estrategias, así como la identificación de información viral en nichos alejados de los denominados "grupos públicos".

Cabe mencionar que es la combinación de estos diferentes aspectos lo que hace problemático el uso de estas aplicaciones, y no características como el diseño privado, el cifrado o la ausencia de filtros. Varias plataformas tienen aplicaciones de diseño privado y no han desempeñado un papel destacado en las últimas elecciones, como Messenger de Facebook, Direct de Instagram, Messages de Twitter, a pesar de su integración con estas plataformas.

6

CONCLUSIONES Y RECOMENDACIONES

El análisis de los instrumentos legales para combatir la desinformación, ya aprobados o en etapa avanzada en el proceso de toma de decisiones, ayuda a iluminar los desafíos de brindar respuestas reguladoras al problema. En primer lugar, por la amenaza que generan a los derechos fundamentales. Hay fuertes críticas dirigidas a casi o todo el contenido normativo de las legislaciones. Asimismo, por diversas razones, en muchos países los artículos han sido posteriormente declarados inconstitucionales por ser abusivos y atentar contra la libertad de expresión.

Es posible situar el principal desafío en dos nodos críticos: el primero es definir y arbitrar sobre 'la verdad'. Toda la tradición reguladora de la comunicación del siglo XX evitó este camino y a ello contribuyó la centralidad que el periodismo profesional tenía en la esfera pública. Con la reorganización del entorno informativo se hizo necesario lidiar con la escala y velocidad de circulación de noticias falsas o engañosas, lo que dificulta que todos los conflictos sean llevados ante la justicia. La falta de parámetros en casi todas las leyes y códigos estudiados para orientar la comprensión de lo que puede considerarse contenido nocivo para las democracias (entre *fake news*, declaraciones falsas, información engañosa en general) es una de las principales razones que dificultan juzgar y regular.

El segundo nodo crítico más importante es la responsabilidad de los intermediarios. El modelo que exime a las plataformas y aplicaciones de los contenidos publicados por terceros ha creado un incentivo para que las empresas no sean lo suficientemente diligentes para abordar los graves problemas de los discursos de odio y la desinformación. Al mismo tiempo, no está claro hasta qué punto es posible pasar de este modelo a otro sin afectar negativamente la libertad de expresión de los usuarios. El camino adoptado por Alemania y el Reino Unido, en el que la plataforma no es responsable de cada contenido individual sino de procesos de cuidado y protección de los deberes de los usuarios parece ser el más prometedor.

En cualquier caso, aún es demasiado pronto para afirmar con certeza que la legislación alemana ha logrado alcanzar este equilibrio. El *White paper* del Reino Unido, por su parte, ni siquiera se ha convertido en ley. El Reino Unido, de hecho, decidió iniciar sus esfuerzos reguladores con mecanismos anticoncentración, que serán objeto de una Unidad de Mercado Digital a partir de abril de 2021.

Hasta el momento no ha habido un debate maduro y buenas prácticas consolidadas. Los consensos que existen giran en torno a asuntos obvios, como la transparencia o los códigos de conducta, que son insuficientes para afrontar la gravedad de la desinformación que ha alcanzado el asunto.

Otro desafío evidente es el intento de abordar la "desinformación" o las "*fake news*" o la "información engañosa" como un solo asunto, cuando en realidad es multifacético. La mayoría de los casos estudiados tienen que ver con centrarse en lo que puede ser explícitamente dañino e ilegal, pero la desinformación puede tomar formas muy diferentes si se conecta a procesos electorales o al discurso de odio contra las minorías, por ejemplo. Las regulaciones que cubren todos los asuntos, políticos o no, y buscan proteger las instituciones gubernamentales y los símbolos nacionales, son las menos democráticas, como las de India y Singapur.

Los procesos de elaboración de normas, ante el reto planteado, han demostrado ser muy distintos. Mientras el Reino Unido hace una consulta pública del documento de definición de escenarios, dilemas y desafíos –y no la propuesta legislativa–, Alemania aprobó la NetzDG en pocos meses y Brasil puede aprobar una ley para combatir *fake news* sin consulta o audiencia pública y durante el periodo de funcionamiento bajo un régimen excepcional del Congreso, debido a las medidas para contener la pandemia del nuevo coronavirus.

El análisis de los casos revela los retos intrínsecos al proceso de construcción de soluciones normativas que sean a la vez protectoras de los derechos y eficaces para promover el acceso a información confiable. Sin embargo, estos desafíos no deben considerarse insuperables. Es necesario ampliar los debates públicos y los esfuerzos de formulación y apostar por soluciones de carácter experimental e innovador, que sean capaces de dar respuestas que impidan que los procesos de desinformación sigan afectando la protección de los derechos y el funcionamiento de las democracias en el mundo.

REFERENCIAS BIBLIOGRÁFICAS

Adams, Paul. Geographies of Media and Communication: A Critical Introduction. Wiley-Blackwell, 2009.

Alemania. NetzDG. Network Enforcement Act, NetzDG. Bonn: 2017. Disponible en: <https://www.bmjv.de/DE/ Themen / FokusThemen / NetzDG / NetzDG_EN_node.html>

----------. Declaration on freedom of expression. [s. d.]. Disponible en: <https: // deklaration-fuer-meinungsfreiheit. desde / en />. Acceso el 28 noviembre de 2020.

Allen, J.; Flores, N. The role of government in the Internet. p. 105, [s. d.]. BfJ - Federal Office of Justice Issues Fine against Facebook. Disponible en: <https://perma.cc/9G3V-SJRN>. Acceso el 26 de noviembre de 2020.

Artículo 19. The Global Expression Report 2019/2020: The state of freedom of expression around the world. Londres, 2020a. Disponible en: <https://artigo19.org/wp-content/ blogs.dir / 24 / files /2020/10/GxR_Final_DigitalVersion_19Oct2020.pdf>. Acceso el 22 de febrero de 2020.

----------. Francia: Analysis of draft hate speech bill. Disponible en: <https://www.article19.org/resources/france-analysis-of-draft-hate-speech-bill/>. Acceso el 24 de noviembre de 2020.

----------. Francia: Avia law is threat to online speech. Disponible en: <https://www.article19.org/resources/ france-avia-law-is-risk-to-online-speech />. Acceso el 24 de noviembre de 2020.

----------. Análisis jurídico de la legislación alemana NetzDG [s. d.]. Disponible en: <https://www.article19.org/wp-content/ uploads / 2017/09/170901-Legal-Analysis-German-NetzDG-Act.pdf>. Acceso el 6 de diciembre de 2020.

----------. Germany: Responding to 'hate speech'. Disponible en: <https://www.article19.org/resources/germany-responding-to-hate-speech/>. Acceso el 6 de diciembre de 2020.

----------. GxR_Final_DigitalVersion_19Oct2020.pdf. [s.d.]. Disponible en: <https://artigo19.org/wp-content/ blogs.dir / 24 / files /2020/10/GxR_Final_DigitalVersion_19Oct2020.pdf>. Acceso el 25 de noviembre de 2020.

Assange, Julian, Appelbaum, Jacob, Müller-Maguhn, Andy, Zimmermann, Jéréme. Cypherpunks: liberdade e o futuro da internet. São Paulo: Ed. Boitempo, 2013.

Bimber, Bruce, Flanagin, Andrew J., Stohl, Cynthia. Collective Action in Organizations: Interaction and Engagement in an Era of Technological Change. Cambridge University Press, 2012.

Brasil. Proyecto de ley 2630/2020. Brasília: Senado Federal, 1992. Disponible en: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2256735>

Breeden, A. French Court Strikes Down Most of Online Hate Speech Law. The New York Times, 18 de junio de 2020.

Bundesamt fur Justiz. BfJ - Federal Office of Justice Issues Fine against Facebook. Disponible en: <https://www.bundesjustizamt.de/DE/Presse/Archiv/2019/20190702_EN.html;jsessionid=306BFD593DD710232937717A-8D07F115.2_ cid393?nn = 3449818 >.> Consultado el 22 de febrero 2021.

Campos, 2020. Lei alemã ou movimento global? O debate sobre regulação de redes contextualizado. Consultor Jurídico. Disponible en: <https://www.conjur.com.

br/ 2020-nov-24 / digital-law-german-law-or-global-movement-contextualizing-debate-Regulation-networks>. Acceso el 22 de febrero 2021.

Canadá. Elections Modernization Act (BILL C-76). Ottawa: Cámara de los Comunes y Senado, 13 de diciembre de 2018. Disponible en: <https://parl.ca/DocumentViewer/en/42-1/ bill/C-76/royal-assent>. Acceso el 22 de febrero de 2021.

CDT. Overview of the NetzDG Network Enforcement Law. Center for Democracy and Technology, [s.d.]. Disponible en: <https://cdt.org/insights/overview-of-the-netzdg-network-enforcement-law/>. Acceso el 26 de noviembre de 2020.

CEPS. The Impact of the German NetzDG law. CEPS, 2 de agosto. 2019. Disponible en: <https://www.ceps.eu/ceps-projects/ the-impact-of-the-german-netzdg-law />. Acceso el 26 de noviembre de 2020.

Christchurch Call. Disponible en: <https: // www. christchurchcall.com/call.html>. Consultado el 22 de febrero de 2021.

CNCDH. Disponible en: <https://perma.cc/HR2L-GH9Q>. Consultado el 24 de noviembre de 2020.

Comisión Europea. Comunicación: Tackling online disinformation: a European Approach. Bruselas: Comisión Europea, 2018. Disponible en: <https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach>

Conseil constitutionnel, 2020. Decisión n° 2020-801 DC du 18 juin 2020 - Communiqué de presse. Disponible en: <https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2020-801-dc-du-18-juin-2020-communique-de-presse>. Acceso el 22 de febrero de 2021.

Corte IDH. OC-5/85. Parecer consultivo del 13/11/85 sobre el registro profesional obligatorio de periodistas. San José: 1985.

Echikson, William, Knodt, Olivia. Conter Extremism Project. Germany's NetzDG: A key test for combatting online hate. Disponible en: <https://www.counterextremism. com/sites/default/files/CEP-CEPS_Germany%27s%20NetzDG_020119.pdf>. Acceso el 26 de noviembre de 2020.

Echikson, W., Knodt, O. Germany's NetzDG: A Key Test for Combating Online Hate. Rochester, NY: Social Science Research Network, 22 nov. 2018a. Disponible en: <https://papers.ssrn.com/abstract=3300636>. Acceso el 24 de noviembre de 2020.

Eletronic Frontier Foundation. Community Input on Christchurch call. 2019. Disponible en: <https: // www. eff.org/files/2019/05/16/community_input_on_christchurch_call.pdf>. Acceso el 24 de noviembre de 2020.

EPRA. Online hate in France - the 'Avia Law': the end of an intensive legislative saga. Disponible en: <https: // www. epra.org/news_items/online-hate-in-france-the-law-avia-an-intensive-legislative-saga-for-a-heantly-censored-law>. Acceso el 24 de noviembre de 2020.

European Commission. Law aimed at combating hate content on the internet. Disponible en: <https://ec.europa. eu/growth/tools-databases/tris/en/index.cfm/search/?trisaction=search.detail&year=2019&num=412&mLang=EN>. Acceso el 24 de noviembre de 2020.

Francia. Loi n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information. Paris: 2018. Disponible en: <https://www.legifrance.gouv.fr/jorf/ id / JORFTEXT000037847559>

----------. Decisión n° 2020-801 DC de 18 de junio de 2020 - Comunicado de prensa del Consejo Constitucional. Disponible en: <https://www.conseil-constitutionnel.fr/actualites/communique/decision-n-2020-801-dc-du-18-juin-2020-communique-de-presse>. Acceso el 24 de noviembre de 2020.

Germany's Balancing Act: Fighting online hate while protecting free speech. Disponible en: <https://www.politico. eu/article/germany-hate-speech-internet-netzdg-controversial-legislation/>. Acceso el 26 de noviembre de 2020.

Global Internet Forum to Counter Terrorism. June 2020 | Appointment of Executive Director and Formation of the Independent Advisory Committee. Disponible en: <https://gifct.org/about/story/#june-2020---appoint-

ment-of-executive-director-and-formation-of-the-independent-advisory-committee-1>. Acceso el 22 de febrero de 2021.

Goldsmith, J., Wu, T. Who Controls the Internet? Illusions of a Borderless World.

Howard, Philip. N. New Media Campaigns and the Managed Citizen. New York: Cambridge University Press, 2006.

ICYMI. New Report on Germany's NetzDG Online Hate Speech Law Shows No Threat of Over-Blocking. Disponible en: <https://www.counterextremism.com/press/icymi-new-report-germany%E2%80%99s-netzdg-online-hate-speech-law-shows-no-threat-over-blocking>. Acceso el 26 de noviembre de 2020.

India. Draft of The Information Technology [Intermediary Guidelines (Amendment) Rules] 2018. New Delhi: MEITY, 2018. Disponible en: <https://www.meity.gov.in/writereaddata/files/Draft_Intermediary_Amendment_24122018. pdf>

Internet Society. The Internet and extra territorial application of laws. 2018. Disponible en: <https://www.internetsociety.org/wp-content/uploads/2018/10/The-Internet-and-extra-territorial-application-of-laws.pdf>. Acceso el 1 de diciembre de 2020.

Jacob, Lisa. Dalia Research. 87% of Germans Approve pf Social Media Regularion Law. Disponible en: <https://daliaresearch.com/blog/blog-germans-approve-of-social-media-regulation-law/>. Acceso el 22 de febrero de 2021.

La Quadrature du Net. Loi haine : le Conseil constitutionnel refuse la censure sans juge. Disponible en: <https://www.laquadrature.net/2020/06/18/loi-haine-le-conseil-constitutionnel-refuse-la-censure-sans-juge/>. Acceso el 24 de noviembre de 2020.

Lasswell, Harold. Propaganda technique in the world war. Peter Smith: Nueva York, 1938.

Lessig, Lawrence. Code: version 2.0. New York: Soho Books, 2006.

Margetts, Helen, John, Peter, Hale, Scott A., Yasse Ri, Taha. Political Turbulence: How Social Media Shape Collective Action. Princeton e Oxford: Princeton University Press, 2016.

Pariser, Eli. The filter bubble. New York: Penguin Books, 2011.

Popkin, Samuel. The Reasoning Voter: Communication and Persuasion in Presidential Campaigns. University of Chicago Press, 1994.

Reino Unido. Online Harms White Paper. Disponible en: <https://assets.publishing.service.gov.uk/government / uploads / system / uploads / attach_data / file / 793360 / Online_Harms_White_Paper.pdf>. Acceso el 22 de febrero de 2021.

Release, P. Victory! French High Court Rules That Most of Hate Speech Bill Would Undermine Free Expression. Disponible en: <https://www.eff.org/press/releases/victory-french-high-court-rules-most-hate-speech-bill-would-undermine-free-expression>. Acceso el 24 de noviembre de 2020.

Reporteros sin Fronteras. Clasificación mundial de la libertad de prensa 2020. Disponible en: <https://rsf. org/pt/classificacao%20>. Acceso el 25 de noviembre de 2020.

Schulz, Jacob. What's Going on With France's Online Hate Speech Law? Disponible en: <https://www.lawfareblog. com/whats-going-frances-online-hate-speech-law>. Acceso el 24 de noviembre de 2020.

Stolton, S. EU Commission to introduce sanctions regime for illegal content in Digital Services Actwww. euractiv.com, 4 de noviembre de 2020. Disponible en: <https://www.euractiv.com/ section/digital/news/eu-commission-to-introduce-sanctions-regime-for-illegal-content-in-digital-services-act/>. Acceso el 24 de noviembre de 2020.

Taylor Wessing. New law to fight online hate speech (Avia law) to reshape notice, take down and liability rules in France. Disponible en: <https://www.taylorwessing.com/en/ insights-and-events/insights/2020/05/new-law-to-fight-online-hate-speech-in-france>. Acceso el 24 de noviembre de 2020.

Tworek, H., Leerssen, P. Transatlantic Working Group. An Analysis of Germany's NetzDG Law. Disponible en:

<https: // www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf. Acceso el 22 de febrero de 2021.

Unión Europea. Código de Prática sobre Desinformação. Bruselas: UE, 2018. Disponible en: <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>. Acceso el 21 de febrero de 2021.

Universal Rights Group. France's watered-down anti-hate speech law enters into force. Universal Rights Group, 16 jul. 2020. Disponible en: <https://www.universal-rights. org/blog/frances-watered-down-anti-hate-speech-law-enters-into-force/>. Acceso el 24 de noviembre de 2020.

Zipursky, R. Nuts About NETZ: The Network Enforcement Act and Freedom of Expression, p. 51 [s. d.].

ACERCA DE LOS AUTORES

João Brant. Investigador y consultor en políticas de comunicación. Doctor en ciencias políticas (USP).

João Guilherme Bastos dos Santos. Investigador del Instituto Nacional de Ciencia y Tecnología en Democracia Digital (INCT.DD). Doctor en comunicación (UERJ).

Tatiana Dourado. Investigadora de la Dirección de Análisis de Políticas Públicas de la FGV y miembro del INCT.DD. Doctora en comunicación (UFBA).

Marina Pita. Posgraduada en derecho digital (UERJ) y estudiante de maestría en comunicación (UnB).

PIE DE IMPRENTA

Friedrich-Ebert-Stiftung (FES)
Calle 71 nº 11-90 | Bogotá-Colombia

Responsables
Omar Rincón
Director de FES Comunicación
omar.rincon@fescol.org.co

Daniela Bohórquez
Gestora de publicaciones
dbohorquez@fescol.org.co

Bogotá, junio de 2021

SOBRE ESTE PROYECTO

FES Comunicación es una unidad regional de análisis de la comunicación para América Latina de la Friedrich-Ebert-Stiftung.

Su objetivo es producir conocimiento para hacer de la comunicación una estrategia fundamental del diálogo político y la profundización de la democracia social. El conocimiento y la red de expertos de FES Comunicación apoyan el trabajo sociopolítico de la red de oficinas FES en América Latina y El Caribe.

Para más información:
https://www.fesmedia-latin-america.org
Facebook @fescomunica
Twitter @fescomunica

REGULACIÓN PARA COMBATIR LA DESINFORMACIÓN.

ESTUDIO DE OCHO CASOS INTERNACIONALES Y RECOMENDACIONES PARA UN ENFOQUE DEMOCRÁTICO

La Unesco y el PNUD definieron la práctica de la desinformación como "contenido falso, manipulado o engañoso, creado y difundido, intencionalmente o no, y que puede causar daños potenciales a la paz, a los derechos humanos y al desarrollo sostenible". De hecho, sus efectos se sienten en muchos campos: en campañas electorales, en asuntos de salud pública (como fue evidente en la pandemia covid-19), en la difusión de discursos de odio contra grupos sociales o en el ataque a la reputación de activistas, y en todas las disputas relevantes en el campo socioambiental, por mencionar los ejemplos más evidentes.

Los ocho casos analizados tienen diferentes objetos de regulación. Los textos que más restringen la libertad de expresión suelen centrarse en la idea de información, declaración o hecho falso. Ninguno de ellos, sin embargo, define criterios públicos para la moderación de contenidos, que es responsabilidad de las empresas de tecnología o responsabilidad directa de los gobiernos. Los sistemas analizados combinan la autorregulación, la corregulación y la regulación pública, pero todavía no hay un debate maduro ni mejores prácticas consolidadas. Los consensos que existen giran en torno a temas obvios, como la transparencia o los códigos de conducta, insuficientes para afrontar la desinformación con la gravedad que ha alcanzado.

Mientras que la protección de las plataformas frente a la responsabilidad por los contenidos de terceros es importante para evitar el efecto silenciador, la cuestión de la responsabilidad de los individuos por los contenidos difundidos en la red está lejos de haber encontrado un punto de equilibrio. Análisis críticos acerca de la respuesta a contenidos de odio, por ejemplo, señalan la baja capacidad de los Estados y los sistemas de justicia para castigar a los autores de delitos relacionados con calumnias, injurias, difamación, racismo, homofobia, transfobia, amenazas, etcétera. En este sentido, sigue siendo pertinente la búsqueda de modelos que puedan proteger la libertad de expresión y, al mismo tiempo, ofezcan mecanismos eficaces para defender los demás derechos humanos con los que pueda entrar en colisión.

FRIEDRICH EBERT STIFTUNG