

INTELIGENCIA ARTIFICIAL, FALSAS NOTICIAS Y DISCURSOS DE ODIOS: ENTRE LA LIBERTAD DE EXPRESIÓN Y LA PERSECUCIÓN PENAL¹ y ²

Marcelo A. Riquert³

ABSTRACT

No hay dudas, Internet tiene un enorme potencial para expandir la libertad de expresión en su doble dimensión: como derecho de todas las personas a difundir e intercambiar ideas y como derecho a buscar y recibir información de todo tipo. Por otra parte, la programación de algoritmos automatizados que potencien la reproducción de falsas noticias y discursos de odio es una cuestión problemática de rigurosa actualidad inserta en aquel contexto. La necesidad y/o conveniencia de su restricción y/o prohibición en determinadas condiciones viene impuesta desde el nivel convencional internacional protector de los derechos humanos. Es una situación de gran complejidad porque lo

¹ **Como citar este artículo científico.** RIQUEERT, Marcelo A. Inteligencia artificial, falsas noticias y discursos de odio: entre la libertad de expresión y la persecución penal. In: **Revista Amagis Jurídica**, Ed. Associação dos Magistrados Mineiros, Belo Horizonte, v. 15, n. 1, p. 221-244, jan.-abr. 2023.

² El presente trabajo fue preparado en el marco del Proyecto del Ministerio de Ciencia e Innovación (España) que lleva por título “La tutela penal de las personas vulnerables: análisis de realidades criminológicas y propuestas sustantivas de ‘lege data’ y de ‘lege ferenda’.” (REF: PID2020-116407RB-I00), bajo dirección del Prof. Dr. Norberto de la Mata Barranco y la Prof.^a Dr.^a Ana Isabel Pérez Machío, ambos de la Universidad del País Vasco (UPV).

³ Doctor en Derecho, UNMDP. Máster de Derecho Penal, U. Salamanca. Director del Área Departamental Penal y de la carrera de posgrado “Especialización en Derecho Penal”, UNMDP. Ex Presidente de la Asociación Argentina de Profesores de Derecho Penal. *e-mail*: riquertm@hotmail.com

que está en tensión es la libertad de expresión, garantizada a nivel constitucional y considerada en la perspectiva cultural de occidente como piedra basal de la democracia.

Palabras clave: inteligencia artificial – falsas noticias – discursos de odio – libertad de expresión – algoritmos – discriminación – derecho penal

SUMÁRIO. 1 Introducción. 2 Las “fake news”. 3 Los discursos de odio. 4 Referencia Normativa. 5 *Fake news*, discursos de odio y libertad de expresión. 6 IA al servicio de la detección/prevenición y de la reproducción/multiplicación. 7 Colofón. Bibliografía y material consultado.

1 INTRODUCCIÓN

Es usual hoy el llamado de atención sobre la proliferación de discursos de odio y falsas noticias potenciadas por algoritmos automatizados. La complejidad para afrontarlo deriva de la “tensión” con la libertad de expresión, constitucionalmente garantizada (arts. 14, 32 y 75 inc. 22, CN; 19 de la DUDH, 4 de la DADH, 19 de la PIDCyP, 13 y 14 de la CDN, 4 de la CIEFDR, III de la CPySDG). No sólo eso, sino que también han sido puestos en foco desde la perspectiva de la preocupación internacional plasmada en diversas convenciones para luchar contra los fenómenos del racismo, la discriminación y la intolerancia.

En el primer sentido, hay consenso sobre que internet tiene una enorme capacidad para expandir aquella libertad en su doble dimensión: como derecho de todas las personas a difundir e intercambiar ideas y como derecho a buscar y recibir información de todo tipo. Así, se entiende que es un instrumento indispensable para el ejercicio pleno de los derechos humanos, siendo el derecho a la libertad de expresión habilitante y posibilitador del ejercicio de otros derechos en internet (lo refleja con sintética elocuencia el conocido

lema: “Internet is free. Internet is freedom. Internet is freedom of speech”) (cf. CHOCARRO, 2017, p. 23). No se profundizará sobre las múltiples consideraciones e implicancias de la libertad de expresión.

En el segundo sentido, luce pertinente recordar que hay límites como la prohibición de propaganda a favor de la guerra o la apología del odio nacional, racial o religioso (arts. 20 del Pacto Internacional de Derechos Civiles y Políticos, 4 de la Convención Internacional sobre la Eliminación de toda forma de Discriminación Racial⁴, 3 de la Convención para la Prevención y la Sanción del delito de Genocidio⁵, 13 inc. 5 de la CADH⁶ y un sinnúmero de declaraciones, propuestas de reglas, etc., de menor jerarquía –soft law–) lo que, claramente, introduce como problema adicional la cuestión atinente al modo de determinar cuándo nos encontramos ante una expresión comprendida en una prohibición o se trata de una admisible exteriorización simbólica de disidencia política o de opinión (por ejemplo, no saludar la bandera o quemarla,

⁴ El citado artículo 4 obliga a los Estados a condenar toda la propaganda y todas las organizaciones que se inspiren en ideas o teorías basadas en la superioridad de una raza o de un grupo de personas de un determinado color u origen étnico, o que pretendan justificar o promover el odio y la discriminación racial, cualquiera que sea su forma, y a tomar medidas inmediatas y positivas destinadas a eliminar toda incitación o actos de tal discriminación.

⁵ El citado artículo 3 indica que, entre otros, serán castigados los actos de instigación directa y pública a cometer genocidio (inc. c).

⁶ El conocido “Pacto de San José de Costa Rica” protege la libertad de pensamiento y de expresión en su art. 13, y su último inciso dice: “Estará prohibida por la ley toda propaganda en favor de la guerra y toda apología del odio nacional, racial o religioso que constituyan incitaciones a la violencia o cualquier otra acción ilegal similar contra cualquier persona o grupo de personas, por ningún motivo, inclusive los de raza, color, religión, idioma u origen nacional”.

En el ámbito americano, no debe soslayarse la importancia que, en la lucha contra la discriminación específicamente de la mujer, tiene la llamada “Convención de Belem do Pará”, vigente desde 1995, así como tampoco la regente desde 2017 y de tono generalista “Convención Interamericana contra toda forma de Discriminación e Intolerancia”.

Si bien adoptada en 2013, aún no vigente, debe recordarse también a la “Convención Interamericana contra el Racismo, la Discriminación Racial y Formas Conexas de Intolerancia”.

llevar vestimentas controversiales en un determinado contexto o con inscripciones que exponen la discrepancia). Vinculado al citado art. 20 del PIDCyP (cuyo 2.º párrafo reza: “Toda apología del odio nacional, racial o religioso que constituya incitación a la discriminación, la hostilidad o la violencia estará prohibida por la ley”), la Oficina del Alto Comisionado de Naciones Unidas en Derechos Humanos (ACNUDH), ha indicado expresamente que, al interpretarse tal norma, el umbral debe ser necesariamente alto porque la limitación a la libertad de expresión debe continuar siendo una excepción, fijando a través del “Plan de Acción de Rabat” (A/HRC/22/17/Add.4, apéndice) un estándar de “prueba” de tal umbral para que una declaración sea considerada como delito⁷.

Maldonado Gutiérrez y Knapp Moreno resaltan que el plan de acción de Rabat no prohija una visión punitiva, sino que propone una mirada preventiva del Estado para atajar estos discursos mediante la promoción del diálogo intercultural, las campañas de difusión, las capacitación de las/os funcionarias/os y una regulación que promueva la pluralidad y diversidad en los medios de comunicación (MALDONADO GUTIÉRREZ; KNAPP MORENO, 2017, p. 30). En línea con ello, recuerdan que la ONG “Article 19”⁸, dedicada a la defensa de la libertad de expresión, propone se busque incentivar los contradiscursos que se opongan a los que promocionan el odio y la violencia, enfatizando que “el antídoto contra el odio no es

⁷ Puede consultarse en <<https://www.ohchr.org/es/freedom-of-expression>>. En concreto, el Plan de Acción de Rabat establece un análisis en seis partes que debe ser satisfecho, vinculado al contexto, el/la orador/a, la intención, el contenido y la forma, la extensión del discurso y la probabilidad, incluyendo la inminencia. Se trata, entre otras cosas, de evitar las persecuciones a las minorías mediante el abuso de leyes, jurisprudencia y políticas poco claras. Entre otras cosas, se llama a los líderes políticos y religiosos a abstenerse de usar la incitación al odio, pero recordando a la vez que tienen un papel crucial en denunciar con firmeza y rapidez las expresiones de odio, dejando en claro que la violencia nunca será tolerada como respuesta a la incitación al odio.

⁸ Cf. <<http://www.article19.org>>. La organización ha elaborado los “Principios de Camden sobre la libertad de expresión y la igualdad”, a partir de debates realizados con funcionarios de las Naciones Unidas y expertos académicos y de la sociedad civil.

en primera instancia el castigo, sino un debate público robusto, incluyente y abierto que exhiba estos discursos”.

Por su lado, destacando la importancia de las “Reglas de Rabat”, nos dice Paz Lloria García (2021, p. 40-41) que se trata de normas que permiten establecer límites en relación con la libertad de expresión y diluir la excesiva punición que se vislumbra en las demandas de castigo de los denominados “delitos de odio”, que es distinto de un delito “motivado por el odio”. En cualquier caso, llama la profesora de Valencia a no olvidar que el derecho penal es una herramienta de “ultima ratio”, que no debe criminalizarse en atención a criterios sostenidos por la moral dominante, por lo que el mero discurso del odio debe buscar otras alternativas a la resolución del conflicto cuando es promovido por particulares con poca proyección. Esto puede ir desde ignorar las manifestaciones para evitar el llamado efecto Streisand, hasta denunciar la cuenta en las propias redes sociales por afirmaciones inapropiadas, produciendo un vacío tecnológico que lleve a su borrado. Su propuesta es que se utilicen mecanismos no penales que pudieran ser eficaces sin recortar la libertad de expresión.

Para preservar esta última en Internet se han elaborado seis principios “rectores”: 1) acceso universal; 2) pluralismo y diversidad; 3) igualdad y no discriminación; 4) privacidad; 5) libre y abierta: transparencia y neutralidad en la red; 6) gobernanza multisectorial (cf. CHOCARRO, 2017, p. 24). No debe pasar por alto que, así como es frecuente que los internautas utilicen las redes sociales para compartir información de interés público, tal vez, su uso sea mayor para difamar o agredir a terceros, lo que se facilita porque se ha naturalizado el hacer circular, retransmitir o copiar opiniones, archivos de imágenes o vídeos que pueden ser tildados de injuriosos o discriminatorios. Se trata de una conducta sobre la que pocas veces se recapacita y que encuentra como habitual excusa para legitimarse que, eventualmente, constituye el ejercicio de la libertad de expresión

(cf. ABOSO, 2021, p. 385).

Cierro la introducción resaltando que el aumento de la circulación de discursos de odio en Argentina en el ámbito de la esfera pública ha sido denunciado desde el Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE), de la Facultad de Derecho de la Universidad de Palermo, señalando que la discusión sobre la legalización del aborto, los debates sobre la política migratoria y sobre el desempeño de las fuerzas de seguridad dieron paso a la aparición de expresiones violentas, agresivas o discriminatorias en la conversación pública (así, TORRES; TARICCO, 2019, p. 2).

2 LAS “FAKE NEWS”

De las Heras Vives señala que cuando hablamos de *fake news* debe fijarse el foco en dos cuestiones: 1) que tienen más que ver con el concepto de realidad que con el de verdad, a pesar de que estén imbricados: la realidad es lo que ocurre verdaderamente y la verdad es la existencia real de algo; 2) las *fake news* por sí mismas no son nada: comunicar lo que verdaderamente no ocurre sin ninguna finalidad es una necedad no exenta de responsabilidad (DE LAS HERAS VIVES, 2020, p. 72).

La CEO de la ONG “First Draft”, Claire Wardle, apunta a la existencia de finalidad cuando dice que las *fake news* “son aquellas creadas con fines de lucro u otros beneficios, difundidas a través de programas patrocinados por el Estado o por medio de la distorsión deliberada de los hechos, por grupos con una agenda particular, incluido el deseo de afectar elecciones políticas” (WEIDENSLAUFER, 2019, p. 3). Se incluye al contenido manipulado, impostor, engañoso, de conexión falsa e incluso la sátira y la parodia. Se habla tanto de información errónea como de desinformación, hechos alternativos o directas mentiras y, es claro,

el principal problema derivado de que las noticias falsas proliferen en Internet y redes sociales es que puedan generar confusión en la ciudadanía al punto que altere su percepción de la realidad.

En realidad, lo único novedoso es el aspecto tecnológico. Hemos pasado de los once principios de propaganda de Göebbels (DE LAS HERAS VIVES, 2020, p. 73) a la construcción de una “posverdad”⁹. Es probable que, en la base del problema, facilitándola, esté que lo que nos llega, se lo recibe y reproduce en forma automática, sin pasarlo por el menor filtro de racionalidad (cf. BRODSKY, 2021, p. 59). Tampoco puede soslayarse que los avances en seguridad digital están detrás de la abundancia de imágenes falsas o manipuladas que se propagan con gran velocidad por Internet. Estudios recientes comprueban que la mayoría de la gente no es capaz de detectar fotos manipuladas. Resalta Julián H. Langevin (2020, p. 130), el gran interrogante es “¿Qué sucede si ya no podemos confiar en nuestros ojos o nuestros oídos?”.

A partir de los usos del “big data” el peligro ya no solo son las noticias falsas, sino también las “*Deep Fake News*” (o Deepfake). Cualquiera, en muy poco tiempo, puede ser víctima de la exposición de falsas intervenciones, participaciones o dichos, dotadas de altísima calidad y, por eso, verosimilitud, con las graves derivaciones consecuentes. Y esta tecnología es, además, cada vez más asequible. Por ejemplo, una app como “Wombo.AI”, es gratuita, está disponible tanto para Android como IOS y lo que hace es animar una foto y transformarla en un video, haciendo que la persona retratada cante, se mueva, y su boca vaya siguiendo el audio (SUEIRO, 2019, p. 92).

Reflexiona Sueiro que, en tiempos de revolución digital y posverdad, las falsificaciones profundas pueden exponerse ante jueces afectando el sistema de administración judicial (SUEIRO,

⁹ El término apareció originalmente en inglés (*post-truth*) en 1992; el diccionario de la RAE ha incluido el neologismo “posverdad” definiéndolo como “Distorsión deliberada de una realidad, que manipula creencias y emociones con el fin de influir en la opinión pública y en actitudes sociales”.

2019, p. 92), agravando la percepción de su crisis. La reiteración de falsedades puede constituir un aporte decisivo si se tiene en cuenta, como han demostrado diversos estudios tanto sobre el “efecto de desinformación” como del “efecto ilusorio de verdad”, que en esta época signada por las versiones sesgadas y falsedades multiplicadas por troles y robots permite advertir que las afirmaciones repetidas en forma consistente son más fáciles de procesar y, de revés, se las percibe como más verdaderas que nuevas afirmaciones, lo que desbarata la idea de que el conocimiento logra doblegar las mentiras (MARADEO, 2021, p. 53-54).

3 LOS DISCURSOS DE ODIOS

El derecho de libre expresión fue pensado como una libertad individual, sin embargo, bajo su amparo, en la actualidad ciertos colectivos sociales propagan discursos contra otros, incitando a la violencia, la discriminación y el odio. En general, debido a múltiples razones que van desde la religión, el género, la orientación sexual, la discapacidad, hasta la nacionalidad, aunque el factor común es que, en todos los casos, lo que se busca favorecer es la intolerancia y la hostilidad. Estos discursos son los que se denominan “de odio” (*hate speech*) y su difusión puede darse tanto en forma oral, escrita, audiovisual, en los medios de comunicación o en Internet. Hay en ellos una apelación a factores emocionales más que racionales, resaltando una actitud de superioridad y dominación en procura de silenciar a grupos sociales con los que se mantiene diferencias étnicas, sociales, sexuales o religiosas (términos como xenofobia, homofobia, racismo, discriminación, delimitan distintos discursos de odio existentes) (cf. BUSTOS MARTÍNEZ et al, 2019, p. 26).

La ya mencionada Oficina del Alto Comisionado de Derechos Humanos de la ONU, brinda una definición de “odio” y “hostilidad” señalando que se refieren a sentimientos intensos e irracionales de oprobio, enemistad y desprecio hacia el colectivo objetivo; mientras

“apología” debe ser entendido como la necesidad de una intención para promover públicamente el odio hacia el colectivo objetivo y el término “incitación” se refiere a declaraciones sobre colectivos nacionales, raciales o religiosos que generen un riesgo inminente de discriminación, hostilidad o violencia hacia personas que pertenezcan a esos colectivos.

Sin embargo, no puede soslayarse la existencia de varias dificultades para definir qué se entiende por “discurso de odio”, como se expone con elocuencia en el proyecto “The Dangerous Speech” de la Universidad Carnegie Mellon, donde se plantea que el término “odio” es impreciso, pudiendo presentar niveles de intensidad diferenciados y llevar a variadas consecuencias. Por caso, cuando se dice “discurso de odio”, ¿significa que quien habla odia? ¿que el hablante quiere persuadir a otros para que odien? ¿que quiere que otras personas se sientan odiadas? Se llama la atención, asimismo, sobre la necesidad de que la ley especifique que identidades y grupos se engloba, o si son todos, no debiéndose perder de vista que si es demasiado general puede provocar su uso indebido contra grupos vulnerables o de la oposición cívica o política¹⁰.

Volviendo con Torres y Taricco, proponen para una mejor comprensión el analizar a los discursos de odio como discursos sociales y nos dicen que, en tanto tales, entienden que en su interior se articula una fuerte unidad entre una determinada concepción del mundo y unas normas de conducta conforme a esa concepción, formas que rechazan la diversidad, la diferencia o la disidencia o el ejercicio de la libertad por el otro. De allí que los discursos de odio

en cualquiera de sus niveles, son articulaciones discursivas que intentan impedir en el otro el ejercicio del derecho a la libertad y a la igualdad... en tanto discursos sociales, buscan imponer una forma única de interpretación sobre los acontecimientos, así como

¹⁰ Cf. <<https://dangerousspeech.org/>>. El 19 de abril de 2021 se ha puesto a disposición una interesante guía al respecto, “Dangerous Speech: a practical guide”. Disponible en: <<https://dangerousspeech.org/guide/>>.

una forma correcta, por lo general tradicional, de desenvolverse en la comunidad (TORRES; TARICCO, 2019, p. 4)

Por eso, aclaran, son generalmente dirigidos contra grupos disidentes, vulnerables, migrantes o cualquier persona o grupo que se visualice como amenazante o responsable por la pérdida de un orden político y social que debe reponerse (TORRES; TARICCO, 2019, p. 4).

Göran Rollnert Liern (2019, p. 95) recuerda que el Tribunal Penal Internacional para Ruanda ha señalado que para la delimitación de las conductas constitutivas de discurso del odio se exige la referencia a tres principios centrales que emergen de la jurisprudencia internacional sobre incitación a la discriminación y a la violencia, a saber, los de intencionalidad, contexto y causalidad, entendida como estimación del impacto probable de la expresión en cuestión.

En términos de su posible tipología o clasificación, recuerdan que las citadas “Reglas de Rabat” distinguen tres tipos de discursos: 1) las expresiones que constituyen delito; 2) las expresiones que no son sancionables penalmente pero que podrían justificar un proceso civil o sanciones administrativas; 3) las expresiones que no son legalmente sancionables pero que aún generan preocupación en términos de la tolerancia, el civismo y el respeto de los derechos de los demás. Resaltan que este orden de diferenciaciones en función al daño que producen, permite distinguir ante los discursos de odio una gama de posibles acciones que no se limiten a la restricción, prohibición o penalización (TORRES; TARICCO, 2019, p. 6).

Más allá de la dificultad conceptual producto de la vaguedad y abstracción de los términos de la legislación internacional que podría dar lugar a descalificar como odio “extremo” a lo que pudiera ser un desacuerdo “común” protegido por la libertad de expresión aunque provoque aversión, desagrado o antipatía (cf. ROLLNERT

LIERN, 2019, p. 105), problema con el que habrá de seguirse liando, es un dato objetivo que en Internet y redes sociales los discursos de odio encontraron una manera más efectiva de propagarse, dando lugar a comportamientos que se han denominado “ciberacoso” y “ciberodio”. Una popular red social, Twitter, viene tratando de morigerar los crecientes niveles de agresión que se advierten en su ecosistema y, para hacerlo, ha comenzado a probar algoritmos de IA ya no para analizar lo que se publica, sino lo que estamos a punto de twittear (TORRES, 2021, p. 11). En el fondo, la iniciativa es la implementación de la censura previa y, por lo tanto, un abierto ataque a la libertad de expresión e, indirectamente, al sistema democrático.

No obstante, reflejo de la clásica bifrontalidad de la tecnología, en ocasiones es el vehículo de logros y reafirmaciones de libertades y derechos en términos democratizantes (paradigmático, el caso “Wikileaks”). Internet nos da la posibilidad de comunicarnos de modo que todos seamos autor y público, y que pocas cosas impidan que digamos lo que queremos decir.

En Argentina, en octubre de 2020, fue oficialmente presentado “NODIO: Observatorio de la desinformación y la violencia simbólica en medios y plataformas digitales”, una iniciativa que ha despertado alguna controversia política y cuyo objetivo es proteger a la ciudadanía de las noticias falsas, maliciosas y falacias. Se lo implementa dentro del ámbito de la Defensoría del Público (organismo creado por la ley de Servicios de Comunicación Audiovisual, n.º 26.522), bajo la órbita del Congreso Nacional, para defender la libertad de expresión. Entre sus funciones está recibir y canalizar los reclamos y consultas de las audiencias. No tiene capacidad sancionatoria alguna¹¹.

¹¹ Fuente: <<https://defensadelpublico.gob.ar/nodio-una-iniciativa-de-la-defensoria-para-fortalecer-la-pluralidad-de-voces/>>. Disponible desde el 12/10/2020; consultada el 10/08/2021.

4 REFERENCIA NORMATIVA

Refiere Weidenslaufer, el “Brexit” y otros eventos electorales recientes provocaron una mirada de especial interés sobre la difusión de declaraciones que tienen por objeto desinformar al público. Desde lo normativo, podrían observarse dos modelos para afrontarlo:

- a) bajo un paradigma de libertad de expresión, mediaría por imperio constitucional (por ej., Primera Enmienda, Constitución de USA, que protege la libertad de religión y de expresión) una imposibilidad de establecer una censura previa de los contenidos de Internet (habría derecho a intercambiar libremente ideas y puntos de vista, con independencia de si son controvertidos o falsos), con lo que sólo puede perseguirse la responsabilidad de quien genera daño con la difusión luego de esta (bajo distintas alternativas como daño o agravio por difamación, invasión a la privacidad, infligir intencionadamente angustia emocional o por interferencia dañosa). Los ISP también están protegidos cuando los contenidos son producidos por terceros, es decir, sin intervención en su creación, desarrollo o edición, ni incumplen promesa legal de retiro de material (sección 230 de la Ley Federal de Decencia en las Comunicaciones de 1996, o CDA). Puede acotarse que parece seguir este esquema el art. 13.2 de la CADH, en cuanto dice que la libertad de pensamiento y de expresión “**no puede estar sujeto a previa censura sino a responsabilidades ulteriores**, las que deben estar expresamente fijadas por la ley y ser necesarias para asegurar: a) El respeto a los derechos o a la reputación de los demás. b) La protección de la seguridad nacional, el orden público o la salud o la moral públicas” (el resaltado es personal);
- b) en Europa, la legislación alemana (de junio 2017, ley contra la publicación en redes sociales de discursos de odio,

pornografía infantil, artículos relacionados con el terrorismo e información falsa, conocida como NetzDG y que ha sido modificada en 2020¹²) o francesa (LO n.º 2018-1201 y n.º 2018-1202, de 22 de diciembre de 2018, relativa a la lucha contra la manipulación de la información), fijan especiales obligaciones a las plataformas como Google+, Twitter, YouTube y Facebook. En la primera, criticada por Human Rights Watch por afectar la libertad de expresión y carencia de supervisión judicial, se procura combatir la difusión de diversos tipos de información (discursos de odio, *fake news* y otro tipo de material ilegal, con expresa remisión a delitos previstos en el CP alemán), con penas de orden económico que pueden llegar de € 500.000 a 50.000.000 y tiene por destinatarias a las plataformas que tengan al menos dos millones de usuarios registrados, que deben eliminar en un plazo de 24 horas el material identificado como ilegal e informar periódicamente sobre las medidas de moderación “transparente” adoptadas. En la segunda, se procura evitar se distorsionen los procesos electorales, creando una nueva vía judicial civil de emergencia para detener la difusión de información falsa durante los tres meses anteriores a una elección nacional¹³.

¹² Se trata de un par de revisiones casi simultáneas, una por la que se requiere a las plataformas de redes sociales que envíen el contenido ilegal a la Policía Criminal Federal; la otra, por la que se reforzaron los derechos de los usuarios al requerir a las plataformas que el proceso de denuncia de contenidos ilegales fuera más intuitivo y permitir la apelación de una decisión de eliminar o no una publicación (UNESCO, 2021, p. 4).

¹³ Weidenslaufer (2019, p. 1 y 4/11). Teruel Lozano (2017, p. 22) aunque coincide en que, con relación al reconocimiento de la libertad de expresión en ordenamientos liberal-democráticos, suelen distinguirse dos modelos, el estadounidense y el europeo, señala que podría existir un tercero, al que llama “latino”, propio de países como España e Italia, que entiende sería abierto y personalista, basado en un reconocimiento de la libertad como un derecho de defensa de la propia persona para su libre desarrollo, incluso egoísta, limitada únicamente cuando ésta genera un daño relevante contra un bien jurídico constitucional principalmente de carácter individual, pero al que al mismo tiempo no desconoce que la libertad tiene una importancia para el orden democrático y se erige en un valor objetivo.

Esta diferencia de enfoque es la que llevó a que esté ausente una regulación específica sobre el discurso de odio en el Ciberconvenio de Budapest de 2001, lo que se salvó un par de años después con el Protocolo Adicional de Estrasburgo (cf. ABOSO, 2021, p. 407).

En Argentina, no hay regulación penal expresa en torno a las “fake news” aunque muchas falsedades, respetando los márgenes de la libertad de expresión, pueden ser atrapadas por diversas figuras penales vigentes. Así, en el CP originario, el art. 109 (calumnia), el 110 (injuria –el 115 la proferida en juicio y el 116 las recíprocas–), el 113 (publicación o reproducción de injurias o calumnias inferidas por otro) y el 114 (calumnia o injuria propagada por la prensa). Posterior, la Ley de Protección de Datos Personales n.º 25326/00, introdujo los arts. 117bis y 157 bis que, luego, sufrieron parcial derogación y modificaciones por Ley 26388/08. El primero, pena el proporcionar a sabiendas a un tercero información falsa contenida en un archivo de datos personales, agravado si se deriva perjuicio a alguna persona.

También son típicas falsedades que resultan mediales, como la estafa del art. 172 del CP; el agiotaje del 300 inc. 1, o los balances falsos (ídem. inc. 2). Innecesario profundizar el ejercicio, queda claro que las falsedades, las mentiras no se penan “per se”, pero cuando se las usa con finalidades de consecuencias dañosas, en muchos casos, tienen repercusiones penales. También, que tenemos ciertas lagunas de tipicidad ante cierta fenomenología de raíz tecnológica.

Se ha observado como un problema, siendo que los ISP tienen un rol indiscutible en el circuito de información (y de desinformación), la falta de previsión que permita penar a las personas jurídicas por este tipo de conductas (así, ELDRID, 2020, punto VI. 2). Creo que, sin apartamiento de un derecho penal de “ultima ratio”, se trata de una cuestión sobre la que, con prudencia, se debiera avanzar.

Respecto de los discursos de odio, cobra importancia el “Protocolo adicional” de Estrasburgo (30/01/2003), que se ocupa específicamente de la xenofobia, el racismo, el odio y la discriminación en Internet. Allí, se define qué se entiende por “material racista y xenófobo” (art. 2.º) y se indica la necesidad de adoptar varias tipificaciones relacionadas con esta temática: a) Difusión de material racista y xenófobo mediante sistemas informáticos (art. 3); b) Amenazas con motivación racista y xenófoba (art. 4); c) Insultos con motivación racista y xenófoba (art. 5); d) Negación, minimización burda, aprobación o justificación del genocidio o de crímenes contra la humanidad (art. 6). También tienen relevancia la resolución del Parlamento Europeo, del 14/03/2013, sobre el refuerzo de la lucha contra el racismo, la xenofobia y los delitos motivados por el odio, que establece la necesidad de adoptar una política criminal contra esas manifestaciones de discriminación; y la DM 2008/913/JAI del Consejo, del 28/11/2008, relativa a la lucha contra determinadas formas y manifestaciones del racismo y la xenofobia mediante el derecho penal (así, cf. ABOSO, 2021, p. 412).

En Argentina, la Ley 23592/88, contra los “Actos discriminatorios”, no ha sido actualizada. Su art. 2.º consagra un severizante genérico para cualquier delito cometido con ánimo discriminatorio y el 3.º tipifica la participación en una organización o en la propaganda de tales actos, así como en su aliento e inicio de realización. No puede soslayarse su eventual conexión con otras conductas punibles previstas en el marco de los “Delitos contra el orden público”, como la instigación a cometer delitos (art. 209, CP), la asociación ilícita (210 y 210bis) u organización o pertenencia a agrupaciones permanentes o transitorias para cometer otros atentados al orden público (213 bis), la incitación pública a la violencia colectiva contra grupos de personas o instituciones (212) o la apología del crimen (213). Es claro, la vieja ley sólo cubre parcialmente los requerimientos del Protocolo.

5 FAKE NEWS, DISCURSOS DE ODIOS Y LIBERTAD DE EXPRESIÓN

Corresponde aclarar que, cuando aquí se habla de libertad de expresión, lo es mirando el fenómeno desde la perspectiva cultural del occidente democrático, donde la mirada jurídica tiende a mantener separados los asuntos de Estado y la religión (cctes.: DE LUCA; LUZZA, 2018, p. 51). En este contexto cultural no se discute su carácter de piedra basal de la democracia. Tampoco que esta, a partir del constitucionalismo liberal pos 2.º GM abrió la puerta a una concepción material que, superando la visión meramente formal de ella, planteó el problema de la necesidad de su protección frente a los enemigos de la propia democracia, dando lugar a las nociones de democracia abierta (que no tiene por qué significar “neutra” o nihilista) y democracia militante. Es importante no perder de vista que una democracia abierta y tolerante reconoce las libertades públicas en su mayor extensión (incluyendo las de sus propios enemigos), pero no tiene por qué ser neutral, sino que cuenta con su propio sistema de valores que promociona activamente, no desde la prohibición y la limitación, sino desde el estímulo de políticas de realización de esos valores democráticos (cf. TERUEL LOZANO, 2017, p. 22).

Se sostiene que la libertad de expresión es la condición indispensable para impulsar el desarrollo en todas las áreas de la vida social, como la cultura, religión, filosofía y política. Y no hay dudas que la aparición de Internet ha facilitado y potenciado la libertad de expresión, lo que se sintetiza diciendo que democratiza el acceso al conocimiento y la información.

No puede soslayarse que hay una suerte de doble vía: todos tenemos el derecho a expresarnos libremente y también el de no ser molestados por las propias opiniones; tenemos el derecho a recibir información y diversos pareceres o puntos de vista, a investigar sobre ello y a difundirlos sin límites geográficos por cualquier medio,

pero no es absoluto, sino que pueden fijárseles restricciones por ley que, usualmente, se sistematizan en dos grandes bloques. Uno, el concerniente a la protección del ámbito privado del sujeto (su intimidad, honra, reputación y buen nombre); el otro, en el ámbito de lo público, se refiere a cuestiones relacionadas con la seguridad nacional, el orden público, la salud y moral públicas (ELDRID, 2020, punto II).

Esto último cobró singular dimensión durante la pandemia del COVID-19, que llevó a alertar a la OMS sobre la existencia de una verdadera “*infodemia*”. De hecho, Unesco y la Oficina del Asesor Especial para la Prevención del Genocidio de las N.U. (OSAPG), a solicitud de la Sección de Libertad de Expresión y Seguridad de los Periodistas de la Unesco, por intermedio del “Oxford Internet Institute” de la Universidad de Oxford, publicó en 2021 un documento ya referido con el título “Hacer frente al discurso de odio en las redes sociales: desafíos contemporáneos”, en el marco del proyecto “#CoronavirusFacts: lucha contra la ‘desinfodemia’ en materia de COVID-19 en entornos propensos a conflictos”. Entre otros aspectos, allí se indica que la estrategia y plan de acción de Naciones Unidas identifica una serie de áreas prioritarias para vigilar y analizar el discurso de odio, estimulando que sus distintas Oficinas estén en condiciones de reconocer, vigilar y analizar las tendencias relativas al discurso de odio y recopilen datos sobre ellas. Específico en cuanto al discurso de odio online, alienta que las entidades de N.U. lleven a cabo más investigaciones sobre la relación entre el uso indebido de Internet y las redes sociales para difundir discursos de odio y los factores que llevan a las personas a cometer actos de violencia, señalando los riesgos que comportan las nuevas tecnologías y las plataformas digitales para la propagación de tales discursos. En lo concreto, el documento denuncia que, durante 2020, la pandemia por el coronavirus puso de manifiesto la propagación de discursos de odio en todo el mundo, agravando la intolerancia y la discriminación hacia determinados grupos, desestabilizando sociedades y sistemas

políticos (cf. UNESCO, 2021, p. 2).

Las campañas de desinformación se ven facilitadas porque la masividad de datos puestos en determinadas manos permite concretar verdaderos ejercicios de ingeniería social, viralizando contenidos a medida del destinatario, de modo que la “fake news” le resulte singularmente atractiva, convincente. Valga la remisión al conocido caso “Cambridge Analytica”, paradigmático en cuanto a la advertencia para la percepción pública de que los datos personales obtenidos pueden no solo ser destinados para el marketing comercial. Reflejo de la preocupación institucional, similar a Europa¹⁴, en el ámbito de la OEA, en octubre de 2019 se ha dado a conocer la “Guía para garantizar la libertad de expresión frente a la desinformación deliberada en contextos electorales” (COMISIÓN, 2019).

6 IA AL SERVICIO DE LA DETECCIÓN/PREVENCIÓN Y DE LA REPRODUCCIÓN/MULTIPLICACIÓN

Destaca Danesi, el uso de algoritmos de IA para la detección del contenido agravante en las redes sociales está en la primera línea de defensa para identificar y remover el denominado “hate speech” o cualquier publicación que infrinja la “política de la empresa” o las “normas comunitarias” (DANESI, 2020, p. 78). A la vez, el problema principal es que aquella operación no está exenta de sesgos y, en consecuencia, los algoritmos terminan discriminando y censurando contenido. Se ha revelado que es mucho más probable que los sistemas de IA destinados a detectar contenido abusivo en línea etiqueten los tweets como “ofensivos” si fueron publicados por

¹⁴ Donde, entre otros instrumentos, ha de tenerse en cuenta la Recomendación General N° 15 de la Comisión Europea contra el Racismo y la Intolerancia para la lucha contra el discurso de odio, de 8 de diciembre de 2015, que fija los criterios de valoración en el caso concreto para habilitar el enjuiciamiento, a saber: contexto general en que el discurso tiene lugar, capacidad del emisor del discurso para influir en terceros, naturaleza y fuerza del lenguaje usado, contexto específico de las afirmaciones, el medio usado y la naturaleza de la audiencia.

personas que se identifican como afroamericanas (DANESI, 2020, p. 78).

Como apunta Álvarez Larrondo, el problema en esta era del “big data” está dado por la convicción generalizada de que las computadoras son capaces de procesar miles de datos y actuar en consecuencia realizando dicha tarea siempre de manera justa y objetiva (además de veloz y barata). Esto es así porque nos han convencido, desde el discurso no contrastado, de que al mediar la intervención de una máquina los procesos dejarán de estar guiados por seres humanos con sus prejuicios: las máquinas que procesan números de manera objetiva. Sin embargo, se trata de una premisa falsa que soslaya que todos los sistemas y algoritmos, en un inicio, han sido concebidos por un humano con sus preconceptos y cargas morales. Estos programas no son más que modelos matemáticos, en líneas generales opacos, difíciles de comprender y sus mecanismos resultan invisibles para todos, inclusive para los matemáticos y los ingenieros informáticos que los pergeñan y elaboran (ÁLVAREZ LARRONDO, 2020). De allí que Cathy O’Neil (2018) haya denominado a este tipo de modelos como “Armas de Destrucción Matemática”.

De Luca y Luzzza (2018, p. 55) alertan que diversos gobiernos vienen implementando medidas respecto al fenómeno de las *fake news*, porque lo consideran una amenaza para la democracia. Así, Alemania, Francia, Italia y España, que han ido desde actualizar la legislación hasta implementar canales de denuncia ciudadana. También lo han hecho agencias y organizaciones internacionales (como el ACNUR) (cf. BUSTOS MARTÍNEZ et al, 2019, p. 31). En Argentina, el Ministerio de Justicia y Derechos Humanos alerta sobre las *fake news* y sus probables consecuencias a nivel personal, sugiriendo consejos prácticos para proteger la propia reputación online e informando sobre los lugares donde recabar asesoramiento y denunciar en caso de ser víctima de falsas noticias (ARGENTINA, 2022).

7 COLOFÓN

La programación de algoritmos automatizados que potencien la reproducción de falsas noticias y discursos de odio es una cuestión problemática de rigurosa actualidad. Su complejidad se amplifica porque lo que está en tensión es la libertad de expresión, garantizada a nivel constitucional y considerada en la perspectiva cultural de occidente como piedra basal de la democracia.

Internet tiene un enorme potencial para expandir la libertad de expresión en su doble dimensión: como derecho de todas las personas a difundir e intercambiar ideas y como derecho a buscar y recibir información de todo tipo. Por tal rol trascendente se elaboraron principios “rectores” de la libertad de expresión en Internet, siendo central un monitoreo serio y comprometido sobre su real vigencia¹⁵.

No debe perderse de vista que la intervención penal debe mantener su nota de ser de “ultima ratio”. La libertad de expresión nació en un sentido contramayoritario. Como crudamente dice Teruel Lozano, para dar protección jurídica a los mensajes que pueden ser dañinos, provocadores e hirientes, sin que esto suponga que gocen de una patente de corso absoluta. Sin embargo, que un mensaje deleznable o nocivo goce del amparo de la libertad de expresión no quiere decir que se debe respetarlo o santificarlo, sino que socialmente

¹⁵ Este monitoreo puede realizarse tanto mediante filtros de palabras clave y métodos de colaboración masiva, como por moderadores humanos que revisan el contenido señalado por usuarios como discurso de odio y deciden si lo es o no. El enfoque manual, personal, tiene ventaja cualitativa en la captación de contexto, pero dificultades en cuanto a la eficacia cuantitativa, lo que se logra mediante la implementación de herramientas de aprendizaje automático y técnicas de procesamiento de lenguaje natural. Entre otras herramientas para detectar el discurso de odio se citan: la plataforma keniana “Umati” (sobre la que brindan ampliada información, TORRES; TARICCO, 2019, p. 6-7); procesando datos de foros de Internet y de Twitter desde 2017 está “Hate Sonar”; la Universidad de Berkeley, a través de ADL y D-Lab, desarrolló “Online Hate Index” (OHI); “Hate Meter” detecta discursos de odio antimusulmanes; “MANDOLA” combina en su proceso de selección tanto el análisis natural de sentimientos, el procesamiento de lenguaje natural como las técnicas de aprendizaje automático y profundo (cf. UNESCO, 2021, p. 506).

puede reaccionarse, aunque el amparo constitucional lo inmune de sanciones jurídicas (TERUEL LOZANO, 2017, p. 25). En este sentido cobra interés el recordatorio de reglas como las de “Article 19” (“Principios de Candel”), en cuanto señalan que los Estados emprenderán esfuerzos amplios para combatir los estereotipos individuales y de grupos negativos, así como la discriminación, promoviendo el entendimiento y la interculturalidad a través de la formación pedagógica sobre los valores y principios de los derechos humanos, desde el diseño curricular escolar y para estudiantes de todas las edades (cf. TORRES; TARICCO, 2019, p. 6). Claro que, en ocasiones, la nota de libertad excederá lo admisible en el debate, que no puede ser libérrimo y, adquiriendo relevante dañosidad para bienes jurídicos, dará paso a la limitación discursiva con consecuencias punitivas.

En Argentina, no hay regulación penal expresa en torno a las “fake news”, lo que no quiere decir que todas las falsedades queden impunes. Respetando entonces la libertad de expresión, pueden ser atrapadas por diversas figuras penales vigentes. Debiera actualizarse la legislación en lo referente al Protocolo de Estrasburgo, que se ocupa de la xenofobia, el racismo, el odio y la discriminación en Internet, e indica la necesidad de adoptar varias tipificaciones relacionadas con esta temática.

BIBLIOGRAFÍA Y MATERIAL CONSULTADO

ABOSO, Gustavo Eduardo. Libertad de expresión, discurso de odio y derecho penal en la sociedad de la tecnología: ¿compartir o dar ‘like’ a una publicación en Facebook puede ser difamatorio? En: RIQUERT, Marcelo A. (Director). **Sistema penal e informática**. v. 4. Buenos Aires: Hammurabi, 2021.

ÁLVAREZ LARRONDO, Federico M. El nuevo derecho artificial: desafíos para el derecho en general. En: ÁLVAREZ LARRONDO, Federico M. (Dir.). **Inteligencia artificial y derecho**. Buenos Aires: Hammurabi, 2020.

ARGENTINA. Ministerio de Justicia y Derechos Humanos. ¿Qué hago si publicaron noticias falsas sobre mí? Mayo de 2022. Disponible en: <<https://www.argentina.gob.ar/justicia/convosenlaweb/situaciones/que-hago-si-publicaron-noticias-falsas-sobre-mi/>>.

ARTICLE19. Principios de Camden sobre la libertad de expresión y la igualdad. Disponible en: <<http://www.article19.org>>.

BRODSKY, Mateo. *Fake news* en tiempos de coronavirus. En: RIQUERT, Marcelo A. (Director). **Sistema penal e informática**. v. 4. Buenos Aires: Hammurabi, 2021.

BUSTOS MARTÍNEZ, Laura; DE SANTIAGO ORTEGA, Pedro Pablo; MARTÍNEZ MIRÓ, Miguel Ángel; RENGIFO HIDALGO, Miriam Sofía. Discursos de odio: una epidemia que se propaga en la red. Estado de la cuestión sobre el racismo y la xenofobia en las redes sociales. En: **Mediaciones Sociales**, Madrid, Ed. Complutense, v. 18, p. 25-42, 2019.

CHOCARRO, Silvia (Org.). **Estándares internacionales de libertad de expresión**: guía básica para operadores de justicia en América Latina. Montevideo: CIMA/NED/Relatoría Especial para la Libertad de Expresión de NU, 2017.

COMISIÓN INTERAMERICANA DE DERECHOS HUMANOS; DEPARTAMENTO DE COOPERACIÓN Y OBSERVACIÓN ELECTORAL; ORGANIZACIÓN DE LOS ESTADOS AMERICANOS-OAS (Departamento de Derecho Internacional de la Secretaría General). **Guía para garantizar la libertad de expresión frente a la desinformación deliberada en contextos electorales**. [S. l.]: OAS, Documentos oficiales, Octubre 2019.

DANESI, Cecilia C. Inteligencia artificial y derecho. En: DANESI, Cecilia C. (Dir.). **Inteligencia artificial, tecnologías emergentes y derecho**. v. 1. Buenos Aires: Hammurabi, 2020.

DANGEROUS SPEECH PROJECT. *Dangerous Speech: a practical*

guide. Disponible en: <<https://dangerousspeech.org/guide/>>.

DE LAS HERAS VIVES, Luis. *Las fake news* ante el derecho penal español. En: **Actualidad Jurídica Iberoamericana**, Madrid, Ed. Instituto de Derecho Iberoamericano, n.º 12 bis, p. 70-77, mayo de 2020.

DE LUCA, Javier Augusto; LUZZA, Yamila Yael. *Fake news: cibercriminalidad y libertad de expresión en internet*. En: RIQUERT, Marcelo A. (Director). **Sistema penal e informática**. v. 1. Buenos Aires: Hammurabi, 2018. p. 51-71.

ELDRID, Brenda. *Fake news*, debate público y libertad de expresión: respuestas penales y extrapenales ante la responsabilidad de los intermediarios de servicios de internet. En: **Temas de Derecho Penal y Procesal Penal**, Buenos Aires, Ed. Erreius, agosto 2020.

LANGEVIN, Julián Horacio. **Juicios penales “virtuales”**: desafíos para garantizar un juicio justo. Buenos Aires: Hammurabi, 2020.

LLORIA GARCÍA, Paz. Discurso de odio antigénero. En: **Otrosí. Revista del Colegio de Abogados de Madrid**, Madrid, n. 10, 7.^a época, 2021.

MALDONADO GUTIÉRREZ, Leopoldo Francisco; KNAPP MORENO, Luis Eduardo. Discurso de odio: ¿ocultarlo mediante el castigo o exhibirlo mediante el debate? En: **Revista Dfensor**, México (DF), Ed. Comisión de Derechos Humanos del Distrito Federal, año XV, n.º 2, p. 26-31, febrero de 2017.

MARADEO, Julián. *Fake news: cómo se fabrican en la Argentina y en el mundo*. Buenos Aires: B, Penguin Random, 2021.

NACIONES UNIDAS. Derechos Humanos. Plan de acción de Rabat. Disponible en: <<https://www.ohchr.org/es/freedom-of-expression>>.

O'NEIL, Cathy. **Armas de destrucción matemática: cómo el Big Data aumenta la desigualdad y amenaza la democracia.** Madrid: Capitán Swing, 2018.

ROLLNERT LIERN, Göran. El discurso del odio: una lectura crítica de la regulación internacional. En: **Revista Española de Derecho Constitucional**, Madrid, Ed. Centro de Estudios Políticos y Constitucionales, n. 115, p. 81-109, enero/abril de 2019.

SUEIRO, Carlos Christian. **Vigilancia electrónica y otros modernos medios de prueba.** 2. edición. Buenos Aires: Hammurabi, 2019.

TERUEL LOZANO, Germán M. Libertad de expresión y discursos de odio en países democráticos. En: **Revista Dfensor**, México (DF), ed. Comisión de Derechos Humanos del Distrito Federal, año XV, n.º 2, p. 21-25, febrero de 2017.

TORRES, Ariel. Ahora los algoritmos te censuran. En: **Diario “La Nación”**, Buenos Aires, edición del 15/5/2021, suplemento “Ideas”.

TORRES, Natalia; TARICCO, Víctor. Los discursos de odio como amenaza a los derechos humanos. In: **CELE/UP**, Buenos Aires, Ed. Centro de Estudios en Libertad de Expresión y Acceso a la Información (CELE), Universidad de Palermo, abril de 2019.

UNESCO - ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA. Unesco y la Oficina del Asesor Especial para la Prevención del Genocidio de las N.U. (OSAPG). **Hacer frente al discurso de odio en las redes sociales: desafíos contemporáneos.** Paris: Unesco, 2021.

WEIDENSLAUFER, Christine. La regulación de las *fake news* en el derecho comparado. En: **ATP**, Santiago (Chile), Ed. Biblioteca del Congreso Nacional de Chile, Sup. n.º 11.973, enero de 2019.

Recebido em: 22-12-2022

Aprovado em: 22-2-2023