



Noticias falsas en Internet: acciones y reacciones de tres plataformas

Febrero de 2021

Facultad de Derecho

Centro de Estudios en Libertad
de Expresión y Acceso a la Información

UP
**Universidad
de Palermo**

Noticias falsas en Internet: acciones y reacciones de tres plataformas*

CELE

En los últimos cuatro años la desinformación en línea se ha convertido en una aparentemente novedosa amenaza a las democracias occidentales. Esta amenaza emerge en un período de ansiedad respecto del futuro: los sistemas democráticos parecen estar atravesando un momento de crisis como resultado de la consolidación de opciones políticas que, si no rechazan abiertamente los procedimientos y formalismos democráticos, los erosionan por dentro. Es en ese contexto que el fenómeno de la desinformación debe ser abordado.

En general, la ansiedad respecto de la democracia produce reacciones. En el caso de la desinformación, quienes han reaccionado han sido los estados, los investigadores, la sociedad civil y las grandes plataformas de Internet por dónde la desinformación parece diseminarse. Este documento explora las acciones de éstas últimas ante los reclamos de los primeros, y enmarca la cuestión en un momento clave sobre el futuro de Internet.

En la primera sección se presenta una conceptualización general del fenómeno de la desinformación en Internet y cómo resurgió en la agenda pública. En este sentido, si bien la difusión de información falsa con fines de engañar al público siempre ha sido parte del repertorio de acciones de distintos actores estatales y paraestatales, los esfuerzos por incidir en el debate público de los Estados Unidos durante la campaña de 2016 inauguraron un período en el que esta vieja práctica adquirió nuevas características, que se replicaron y a la vez mutaron en procesos electorales posteriores como los del Reino Unido (2016), Colombia (2016), México (2018) y Brasil (2018). Esta dimensión temporal del fenómeno es relevante para su definición, ya que permite asociar las reacciones de las grandes plataformas de Internet con una crisis puntual, que se alimenta—también—de un extendido giro *pesimista* respecto de la capacidad de Internet de incidir de manera positiva en el futuro de la democracia.

* Este documento fue escrito por Ramiro Álvarez Ugarte y Agustina Del Campo.

La segunda sección presenta un mapa general de las principales acciones de las plataformas en materia de desinformación, desde fines de 2016 hasta fines de 2020. El objetivo de esta aproximación temporal es doble: identificar acciones anunciadas, especialmente en el contexto de América Latina, y constatar el grado de implementación de las mismas. La lenta acumulación de cambios anunciados y respuestas al fenómeno sugieren una industria bajo presión, intentando encontrar respuestas satisfactorias a una creciente demanda de moderación de contenidos problemáticos.

La tercera sección discute los hallazgos y postula la existencia de cierta “aporía” en el centro del problema. El giro *pesimista* respecto de Internet y el proceso de concentración del tráfico en las grandes plataformas han generado nuevos cuestionamientos sobre la solución legal alcanzada hasta el momento respecto del flujo de información en línea. Las acciones de las plataformas sugieren que están avanzando hacia un modelo de *foro público*, donde los principios y criterios que utilizan para definir qué discurso es aceptable y cuál no están enteramente bajo su control. Ese modelo pone en cuestión el principio de *no responsabilidad de intermediarios por los contenidos de terceros* que hasta el momento era la solución legal predominante.

A. La desinformación como problema actual de la democracia

Una definición posible de la desinformación sostiene que se trata de “la difusión masiva de información falsa (a) con la intención de engañar al público y (b) a sabiendas de su falsedad”.¹ Tomar a esta definición como punto de partida nos permite ver que no estamos ante un fenómeno especialmente novedoso: diversos actores han procurado manipular a la población a través de información falsa a lo largo de la historia.² Quizás por ello, para el análisis del fenómeno parece ser más relevante acotarlo temporalmente que conceptualmente.

En efecto, la *desinformación* emerge como fenómeno reciente en el marco de

¹ CIDH («Guía para garantizar la libertad de expresión frente a la desinformación deliberada en contextos electorales». Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos, Washington D.C. 17 de octubre de 2019.), 3. Ver también C. Boterolgnacio Álvarez, Eduardo Bertoni, Catalina Botero, Edison Lanza (eds.) («La Regulación Estatal De Las Llamadas “noticias Falsas” Desde La Perspectiva Del Derecho A La Libertad De Expresión», en *Libertad de expresión: A 30 años de la Opinión Consultiva sobre la colegiación obligatoria de periodistas*, 1a, Comisión Interamericana de Derechos Humanos, Washington D.C., 2017, (OAS. Documentos oficiales ; OEA/Ser.D/XV.18.), 69; M. Verstraete; D. E. Bambauer; J. R. Bambauer («Identifying And Countering Fake News». Social Science Research Network, Rochester, NY. ID 3007971. 1 de agosto de 2017.)

² Varios estudios han señalado este punto: ver L. Bounegru; J. Gray; T. Venturini; M. Mauri («A Field Guide To “Fake News” And Other Information Disorders». Public Data Lab & First Draft, Amsterdam. 2017.), 6; R. Darnton («The True History Of Fake News», THE NEW YORK REVIEW OF BOOKS, 13/2/2017, disponible en <https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/> Fecha de consulta: 1/marzo/2019.)

procesos electorales caracterizados por un debate público empobrecido en parte gracias a la difusión masiva de informaciones falsas, especialmente luego de la campaña electoral de los Estados Unidos de 2016 que llevó a Donald J. Trump a la presidencia de ese país.³ Desde entonces, en diversas elecciones la desinformación apareció como un problema más o menos serio: los referéndum por la permanencia del Reino Unido en la Unión Europea y por la paz en Colombia en 2016 y las elecciones presidenciales de Brasil y México de 2018 son buenos ejemplos de ello.

Diversos observadores concluyen que la desinformación es una amenaza seria para el futuro de los sistemas democráticos, ya que los procesos electorales son el principal mecanismo a través del cual los gobiernos adquieren legitimidad. Esta percepción, compartida por parte de algunos estados, generó reacciones de diverso tipo—regulatorias, de concientización, o de autorregulación—que aumentaron la presión sobre las grandes plataformas de Internet, percibidas como “facilitadoras” del fenómeno.⁴ La presión estatal, sumada a los daños a la reputación producidos por prácticas empresariales cuestionables en el manejo de datos personales de los usuarios, son la principal causa de las acciones anunciadas o adoptadas por las plataformas en los últimos años, que este informe documenta. Pero entender esas reacciones exige comprender la temporalidad del fenómeno. La *desinformación* se inserta en dos crisis distintas pero relacionadas: la crisis de la democracia; y la del modelo *optimista* de Internet que justificó—hasta el momento—las principales aproximaciones regulatorias a la forma en la que circula la información en esta red.

1. Internet y la era del optimismo tecnológico

En tanto red descentralizada, Internet se presentó desde el principio como el espacio ideal para la libre circulación de información e ideas de toda índole.⁵ La creación de la red, y—especialmente—su masificación a partir de la segunda mitad

³ Cf. C. Cortés; L. Isaza, «Noticias falsas en Internet: La estrategia para combatir la desinformación». Centro de Estudios para la Libertad de Expresión, Buenos Aires. diciembre de 2017. Pág. 2 (dónde referencia al famoso «Pizzagate»).

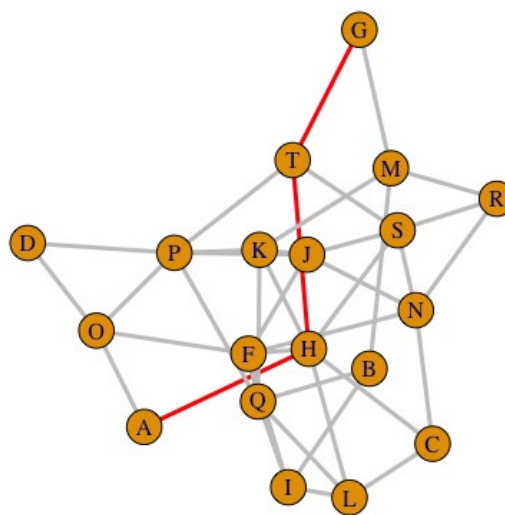
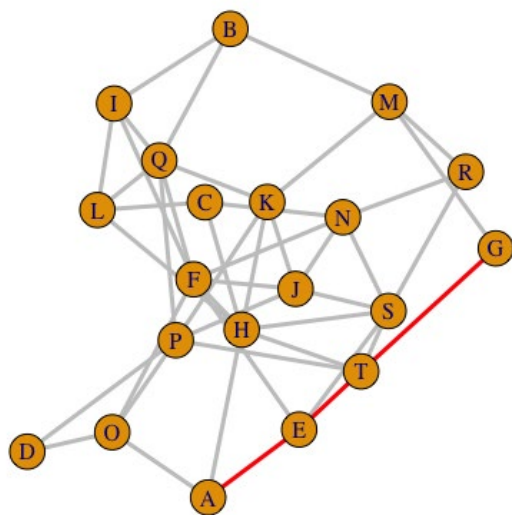
⁴ Dentro de estas reacciones, cabe señalar a la ley alemana del 30 de junio de 2017, conocida como *Netzwerk Durchsetzungsgesetz* o NetzDG, que impone diversas obligaciones a las plataformas en relación a diversos contenidos ilegales. La norma fue criticada por el Relator Especial sobre la Libertad de los Medios de la OSCE y por Human Rights Watch. Cf. OSCE (Osce Representative On Freedom Of The Media Warns Germany Social Networks Law Could Have Disproportionate Effect, OSCE, 04/10/2017, disponible en <https://www.osce.org/fom/347651> Fecha de consulta: 4/marzo/2019.); H. R. W. |. 350. F. Avenue; 34th. F. |. N. York; N. 10118. U. |. t 1.212.290.4700 (Germany: Flawed Social Media Law, HUMAN RIGHTS WATCH, 14/02/2018, disponible en <https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law> Fecha de consulta: 16/enero/2019.). Francia sancionó una norma similar en noviembre de 2018, específicamente dirigida a combatir la “manipulación de la información.”

⁵ Sobre esta descripción, ver p.ej. Y. BENKLER, *The Wealth Of Networks: How Social Production Transforms Markets And Freedom*, Yale University Press, New Haven [Conn.], 2006 (donde se recoge esta visión «optimista» sobre la red).

de la década de 1990, cambió radicalmente el modo en que fluye la información: de modelos altamente centralizados y controlados por actores poderosos se pasó a un modelo descentralizado en donde, con una muy baja inversión en tecnología, cada uno de los usuarios podría producir contenido propio y difundirlo en condiciones similares a las de los grandes productores de contenidos tradicionales.⁶

Esta promesa inicial de descentralización generó cierto optimismo respecto de Internet y su influencia en el futuro de la democracia, porque algunas características básicas del diseño de la red tienden a favorecer el flujo de información por sobre el control del mismo.⁷ Este diseño fortalece la resiliencia de la red y hace

C de control



⁶ Cf. Y. Benkler (*Ibid.*), cap. 5 (donde poner por ejemplo a la famosa producción cinematográfica casera *The Jedi Saga*). Gran parte de la literatura pionera en materia de derecho e Internet señalaba esta dimensión potencialmente beneficiosa de Internet. Ver también p.ej., L. Lessig (*Code*, Version 2.0, Basic Books, New York, 2006.); E. Moglen («The Invisible Barbecue», COLUMBIA LAW REVIEW, vol. 97, 1997.) (“Any individual can reach through the use of network media such as the World Wide Web an audience far larger than the owner of several small television stations, and at no cost. The form of communication achieved is no doubt different, and it is obvious that the individual UseNet news post has not achieved parity with a television advertisement for an athletic shoe. But the metaphor of the broadcaster and the consumer conditions us to accept the maximum possible inequality as a natural necessity: a few talk and the rest merely listen. Spectrum need not be allocated around this conception; the historical justifications for doing so are among the forms of reasoning vitiated by the rise of complementary carriage technologies and the resulting death of ‘scarcity rationales’”)

⁷ Cf. L. Lessig (*Code*, cit.), ix-x; un optimismo que—en ocasiones—derivó en lo que Eugeny Morozov llamó acertadamente cyber-utopianismo. Cf. E. Morozov (*The Net Delusion: The Dark Side Of Internet Freedom*, 1st ed, Public Affairs, New York, 2011.). Ver también J. Zittrain (*The Future Of The Internet And How To Stop It*, Yale University Press, New Haven [Conn.], 2008.), 30-35 (donde describe los principios de diseño de la Internet original)

⁸ El principio de *end-to-end* (e2e) sobre el cual se montó la primera versión de Internet delega la mayor parte del trabajo en los extremos de la red: la red debe ser “lo más simple posible, y su inteligencia debe estar en los extremos o los puntos finales de la red”.[#] Este principio explica por qué al hablar de Internet hablamos de red *descentralizada* — la red podría haber sido diseñada de otra manera, mediante p.ej. el establecimiento de puntos nodales a los que cada usuario deba conectarse para acceder. Pero no fue así.[#]

Figura 1

(a) Red aleatoria descentralizada de 20 nodos

(b) Red aleatoria descentralizada de 19 nodos (-E)

La Figura 1 representa una red descentralizada, en la que los nodos se conectan entre sí de diversas maneras. Una red de esta naturaleza, para ser eficiente, busca que los puntos más alejados entre sí se conecten mediante los pasos más cortos posibles. En la Figura 1.a es posible ver resaltado el vínculo entre el nodo A y el nodo G: el camino más corto posible lleva tres pasos ($A \rightarrow E \rightarrow T \rightarrow G$). Si removemos el primer nodo de ese tramo ("E", en la Figura 1), vemos que el nodo encuentra rápidamente otro camino igual de eficiente ($A \rightarrow H \rightarrow T \rightarrow G$). Una red menos descentralizada, en la que existen nodos que cumplen un rol de *brokers* (como, p.ej., el nodo "K" en la Figura 2), la conexión de los nodos $D \rightarrow G$ se vuelve imposible.

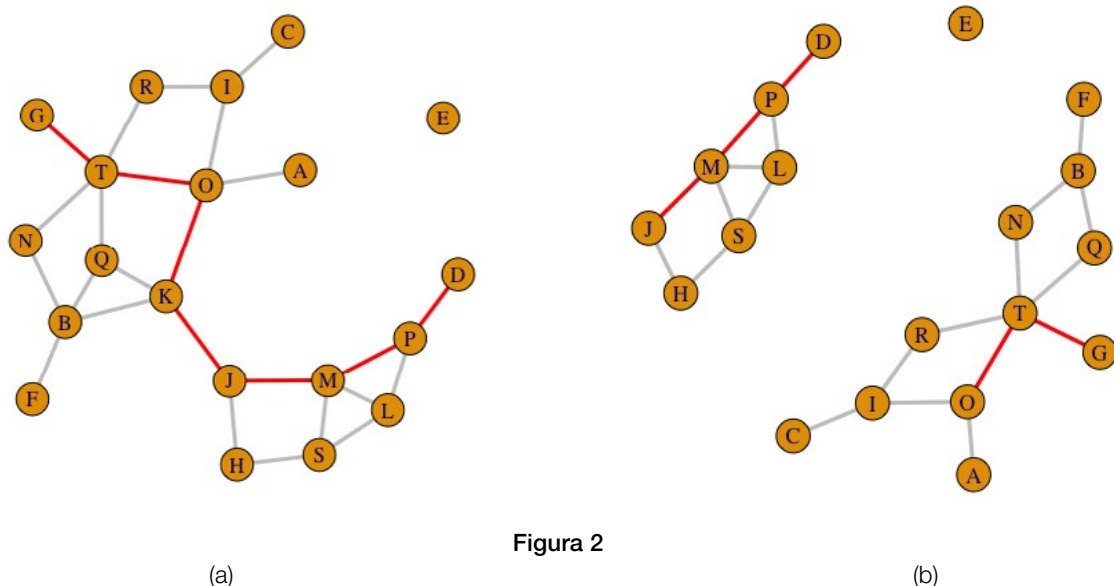


Figura 2

La masificación de Internet trajo aparejado el cuestionamiento de algunas presunciones del modelo original. Así, por caso, hacia fines de la década de 1990 los riesgos para la seguridad que representaba una red descentralizada se volvieron evidentes.⁹ Sin embargo, la versión *optimista* perduró y se expresó en cierta confianza en que Internet podría ser una herramienta democratizadora en sociedades cerradas, ya que su carácter descentralizado permitiría corroer las estructuras que permiten

⁹ Sobre este punto, ver especialmente *Ibid.*, cap. 3.

a los gobiernos totalitarios controlar a sus ciudadanos.¹⁰ En ese contexto, pronto surgieron protecciones legales para los actores intermediarios que facilitaban el acceso a contenidos producidos por terceros. La sección 230 de la *Communications Decency Act* de 1996 estableció, por caso, el principio de *no responsabilidad* de los intermediarios en resguardo—precisamente—del diseño arquitectónico original de la red. ¿Quién se arriesgaría a cumplir ese rol de intermediación si pudiera ser responsabilizado por la información cuya circulación facilita? Este modelo fue luego adoptado en otros países, por vía legislativa, jurisprudencial o de *soft law*.¹¹

2. La era del pesimismo

Las visiones optimistas sobre Internet tuvieron desde el comienzo sus detractores. Numerosos autores vieron cómo ciertas tendencias en el desarrollo de la red podrían derivar en una Internet más cerrada, regulada y sujeta al control de actores estatales y privados. Así, Lawrence Lessig señaló en 1999 cómo las presiones del comercio y los estados tenderían a regular a Internet de una forma mucho más estricta, tanto a través de legislación como de requisitos en el mismo *código* con el que se diseña la arquitectura de la red.¹² En el mismo sentido, Jonathan Zittrain predijo cómo las preocupaciones por la seguridad podrían derivar en el retorno de modelos tecnológicos que Internet parecía haber dejado atrás (los “electrodomésticos” controlados por los fabricantes que hacen sólo lo que los fabricantes desean, p.ej., los teléfonos celulares).¹³ Cierta cautela fue emergiendo: Internet no era una tecnología *necesariamente* beneficiosa para la democracia. Eugeny Morozov fue uno de los primeros en señalar cómo los gobiernos autoritarios estaban aprendiendo a usar Internet para aumentar su control sobre la población.¹⁴ Estas advertencias fueron marcando el tiempo de cierto “giro pesimista”, que estuvo sin dudas alimentado por la mutación de la arquitectura original de la red.

En el proceso de expansión de Internet surgieron diversos actores intermediarios en el flujo de información que se volvieron cada vez más poderosos. Así,

¹⁰ E. MOROZOV, *The net delusion*, cit., *passim*.

¹¹ Ver p.ej., los Manila Principles, en <https://www.manilaprinciples.org/es>. En contra de esta visión, ver J. KOSSEFF, *The Twenty-six Words That Created The Internet*, Cornell University Press, Ithaca [New York], 2019, cap. 7 (dónde señala que la sección 230 es una muestra del «excepcionalismo» de los Estados Unidos, y argumenta que otros países han adoptado versiones más moderadas de esa provisión).

¹² L. LESSIG, *Code*, cit.

¹³ J. ZITTRAIN, *The Future Of The Internet And How To Stop It*, cit., cap. 1.

¹⁴ E. MOROZOV, *The net delusion*, cit.

Google (1998), por ejemplo, logró una posición dominante en el mercado de los motores de búsqueda y se convirtió en un actor fundamental para acceder a información que no sabemos que existe. YouTube (2005) hizo algo parecido con el mercado de los vídeos en línea, de la mano de un aumento de las velocidades de conexión. Facebook (2004) y Twitter (2006) permitieron conectar a las personas replicando—en Internet—las redes sociales físicas y creando nuevas. Estas empresas empezaron a concentrar gran parte del tráfico y así se fue dando un proceso de *re-centralización* de la red.¹⁵ Estos actores se expandieron, adquirieron nuevos servicios—en muchos casos, competidores directos—y consolidaron sus posiciones dominantes gracias a los efectos de red que generan incentivos para participar de esos canales centralizados (y costos crecientes de “desconexión”).¹⁶

En el marco de esa *re-centralización*, estas empresas intermediarias desarrollaron términos de servicios y *community guidelines* que operan como regulaciones *de facto* de la información que es reputada aceptable o no en cada una de estas plataformas. Además, abrazaron de una manera más decidida su rol de moderación cuando introdujeron las *timelines* basadas en algoritmos que—se supone—buscan acercar a los usuarios contenidos más “relevantes” para ellos, determinados por patrones de consumo que las propias plataformas registran y explotan.¹⁷

¹⁵ Ver S. Hubbard («Fake News Is A Real Antitrust Problem», 12/2017, disponible en <https://www.competitionpolicyinternational.com/fake-news-is-a-real-antitrust-problem/> Fecha de consulta: 16/enero/19), 2 (“Two corporations have an outsized control on the flow of information worldwide. Google accounts for roughly 80 percent of global Internet searches, and its search market share exceeds 90 percent in most European countries. Facebook dwarfs all other social networks, with two billion active monthly users”). Ver también Observacom («Ott Regulation. Key Points For The Democratic Regulation Of “over-the-top” Services So As To Ensure A Free And Open Internet And The Full Exercise Of Digital Rights And Freedom Of Expression». Observacom. septiembre de 2017.), 3:

“In a scenario centralized by the traditional media, it was clear that the market on its own did not guarantee the fundamental diversity, pluralism and freedom of expression needed by democracy. With the emergence of the Internet, it seemed that part of the rationality that gave meaning and foundation to democratic regulation might have been lost. In fact, some important players in the digital ecosystem claim that regulation of the Internet is not only dangerous but should not exist, as it is no longer necessary or possible. However, after the initial phase of more decentralized and open network operation, new bottlenecks formed and the Internet embarked on a growing centralization among just a few actors of the digital ecosystem that has affected its potential to serve all of humanity: this was underlined by the creator of the World Wide Web, Tim Berners Lee. The trend towards concentration and threats to freedom of expression on the Internet show that diversity and pluralism—and even the notion of an open and free Internet—need regulatory guarantees so that they can be maintained as values and paradigms of modern digital communications”.

¹⁶ Sobre este punto, ver K. S. Baran; K. J. Fietkiewicz; W. G. Stock («Monopolies On Social Network Services (sns) Markets And Competition Law», ISI, 2015.) (describiendo los efectos de red que llevan a monopolios o cuasi-monopolios en los mercados de redes sociales). Ver también E. Morozov (*The net delusion*, cit.), 217 (“While activists can minimize their exposure to intermediaries by setting up their own independent sites, chances are that their efforts may not receive the kind of global attention that they might on Facebook or YouTube”)

¹⁷ Este modelo es diferente al que estaba basado, pura y exclusivamente, en las elecciones de los usuarios: desde los “portales” que curaban la información manualmente hasta el modelo de “lectores de RSS” que acercaba información fácilmente accesible pero con base en la suscripción “expresa” de los usuarios. Sobre este cambio, ver M. A. DeVito; D. Gergle; J. Bir-

Estos comportamientos pusieron en evidencia el importante rol que cumplen en los hechos, y llevó a muchos legisladores alrededor del mundo a preguntarse por qué los organismos representativos de la democracia no pueden intervenir en la definición de esos criterios.¹⁸ Este poder, que ejercieron erráticamente pero de forma sostenida los intermediarios, con el tiempo se vio acompañado de una creciente demanda de diversos actores por mayor responsabilidad y rendición de cuentas.¹⁹ Y esto, a su vez, ha empujado a las plataformas a hacer mucho más que simplemente “facilitar” el flujo de información: se han convertido en “curadores” de esos contenidos, tanto por razones propias (como el desarrollo de algoritmos de recomendación de contenidos) como por presiones externas.²⁰

Finalmente, cabe destacar que este proceso de concentración fue facilitado por la expansión de un modelo de negocios que da estos servicios de manera “gratuita”, bajo una tradicional lógica publicitaria que se vale—sin embargo—del análisis y explotación de los datos personales de los ciudadanos para dirigir mensajes altamente personalizados.²¹ Este modelo fue objeto de diversas críticas en los últimos años, en parte motorizadas por ciertos escándalos que revelan el nivel de afectación a la privacidad que el modelo supone, entre ellos las revelaciones de Edward Snowden en 2013 o el escándalo de *Cambridge Analytica* y el uso de

nholtz («“Algorithms Ruin Everything”: #riptwitter, Folk Theories, And Resistance To Algorithmic Change In Social Media», CHI '17: CHI Conference on Human Factors in Computing Systems, Denver Colorado USA, 2 de Mayo de 2017, disponible en: <https://dl.acm.org/doi/10.1145/3025453.3025659>)

¹⁸ Sobre este punto, las tendencias son por supuesto contradictorias: mientras algunos quieren endurecer los criterios de las plataformas para lidiar con diversos males (desinformación, difamación, violaciones de derechos de autor, discurso de odio, etcétera) otros quieren resguardar a los usuarios de decisiones arbitrarias o basadas en criterios opacos.

¹⁹ Sobre este punto, la literatura sobre “accountability” y empresas tecnológicas parece haber aumentado exponencialmente en los últimos años. Ver S. Pearson; N. Wainwright («An Interdisciplinary Approach To Accountability For Future Internet Service Provision», INTERNATIONAL JOURNAL OF TRUST MANAGEMENT IN COMPUTING AND COMMUNICATIONS, vol. 1, 1, 2013, Disponible en <https://www.inderscienceonline.com/doi/abs/10.1504/IJTMCC.2013.052524>); R. Gajanayake; R. Iannella; T. Sahama («Sharing With Care: An Information Accountability Perspective», IEEE INTERNET COMPUTING, vol. 15, 4, 2011.)

²⁰ Cf. Observacom («Ott Regulation. Key Points For The Democratic Regulation Of «over-the-top» Services So As To Ensure A Free And Open Internet And The Full Exercise Of Digital Rights And Freedom Of Expression», cit.), 9 («Such intermediaries no longer provide just technical support and ‘transit routes’, but often affect the contents that circulate through such routes. Not only are they able to monitor all content produced by third parties, but they can also intervene in them, ordering and prioritizing their access and, therefore, determining what contents and sources of information a user may or may not view. They can also block, eliminate or de-index content—such as speeches protected by the right to freedom of expression—as well as users’ accounts or profiles. These actions are often forced by external pressures from government authorities or other private actors, but also by the decisions taken by the intermediaries themselves»). Ver también L. Lessig (*Code*, cit.), cap. 6 (sobre la historia de la moderación de contenidos en las primeras “comunidades” de Internet) y S. T. Roberts («Content Moderation», 2017.), 3-4 (llamando la atención sobre una “nueva escala” de la moderación como consecuencia de la concentración, movida en parte por los riesgos “legales, financieros y de reputación” que estas compañías públicas que responden a sus accionistas deben afrontar por el contenido generado por los usuarios sin control alguno).

²¹ Cf. D. Ghosh; B. Scott, «Digital Deceit: The Technologies Behind Precision Propaganda On The Internet». New America, Washington D.C. enero de 2018.

datos personales obtenidos de Facebook con fines políticos de 2018.²² Este modelo publicitario no se encuentra necesariamente en crisis pero las críticas de los últimos años contribuyeron al giro “pesimista”.

3. La desinformación como dilema

Desde la elección presidencial de los Estados Unidos en 2016, las principales plataformas intermediarias parecen estar a la defensiva. De manera regular y un tanto caótica, desde entonces han anunciado e implementado cambios en sus plataformas destinados a combatir el fenómeno de la desinformación. Asimismo, se han visto convocadas a declarar en audiencias informativas ante organismos legislativos de todo el mundo. También han publicado estudios y ofrecido información en el marco de esfuerzos de transparencia destinados a aplacar a sus críticos. Han apoyado, y en algunos casos liderado, programas de fortalecimiento del periodismo de calidad y la “verificación de información” como remedios poco problemáticos.

Estas reacciones no se explican por fuera del giro *pesimista* previamente descrito. Es que lo que se cuestiona no sólo es el rol que tienen en el flujo de información en línea sino también el marco regulatorio que, hasta el momento, determinó las obligaciones y responsabilidades de los intermediarios en Internet. En efecto, el principio de *no responsabilidad de los intermediarios* tenía sentido en una red descentralizada como la representada en la Figura 1: establecer responsabilidades en ese escenario implicaba obstaculizar, en lugar de favorecer, el libre flujo de información. Pero en una red *centralizada*, ¿tiene sentido mantener el mismo principio?

El rol que las empresas asumen en materia de desinformación—en muchos casos, bajo una creciente presión social y política—plantea un dilema más general. Éste último puede presentarse mediante la siguiente pregunta: ¿quién lidiaba con la desinformación antes de Internet? Los estándares de libertad de expresión negaban todo tipo de autoridad al estado para definir qué es verdadero y qué falso: ese juicio quedaba reservado a cada uno de los ciudadanos que, en el marco de un libre mercado de ideas, podían acceder a distintos puntos de vista y formar su propia opinión. Este arreglo no fue—solamente—una apuesta a una ciudadanía virtuosa: se trata de una regla de adjudicación de facultades y funciones en una sociedad democrática. El estado no *debe* intervenir en determinar qué es verdadero y qué es falso, y eso explica por qué la jurisprudencia sobre libertad de expresión del siglo XX fue tan protectora de este derecho alrededor del mundo.

²² Este modelo ya había sido señalado por Lessig. Cf. L. LESSIG, *Code*, cit., cap. 11.

Las respuestas a la desinformación que estamos viendo hoy en día de parte de las plataformas problematizan ese estado de cosas por dos razones. Primero, porque estas empresas son en parte empujadas por reclamos estatales de mayor acción, violando el principio señalado con anterioridad respecto del rol del estado en el discernimiento entre lo verdadero y lo falso. Súbitamente, el estado ha legitimado informalmente—a través de las demandas y presiones de algunos de sus funcionarios—a un actor intermediario, central para el flujo de información, como árbitro de la verdad. Los ciudadanos, según parece, ya no pueden hacer en la era de Internet lo que hacían antes. Segundo, en tanto empresas privadas, toman estas acciones en el marco de un derecho legalmente protegido de raigambre contractual, sin adecuados mecanismos de rendición de cuentas o de responsabilidad. Para decirlo de otro modo, hemos incorporado un actor nuevo que toma decisiones que antes reservábamos a cada individuo y ni siquiera se trata de un actor sobre el que podamos influir mediante procedimientos democráticos. Estos desafíos presentan, entonces, un problema a resolver: ¿mantenemos nuestro compromiso con la libertad de expresión amplia y volvemos a confiar en que la mejor respuesta a la información falsa es la libre competencia del mercado de ideas o lo revisamos? En el segundo caso, ¿con qué lo reemplazamos?

B. Acciones

Para la documentación de las acciones desplegadas por las plataformas en los últimos años usamos cuatro categorías:

1. *Acciones de concientización.* Aquí agrupamos las acciones políticas de las plataformas, incluyendo alianzas con otros actores, la promoción de campañas de educación, alfabetización digital y de medios, etcétera. Son acciones que buscan construir un ecosistema positivo respecto de la desinformación, para empoderar a actores y estrategias que—se supone—pueden combatir al fenómeno. Por definición, no suponen cambios ni en el código ni en las políticas de las plataformas.
2. *Cambios de código.* Aquí nos referimos a las acciones que modifican el código de las plataformas, es decir, el modo en que éstas funcionan. Se incluyen cambios a los algoritmos de recomendación, visibilización y alcance de contenidos.
3. *Cambios de política y acciones de moderación.* Aquí nos referimos a las acciones que modifican las políticas de las compañías, tanto los *Terms of Ser-*

vice (TOS) como las políticas de moderación de contenidos (*community guidelines*). También aquí agrupamos a las acciones que implementan políticas internas o mandatos externos de remoción de contenidos reputados ilícitos o contrarios a las guías de la comunidad, en casos de desinformación.

4. *Acciones de transparencia y relaciones públicas.* Aquí agrupamos las acciones que revelan información sobre el funcionamiento de la plataforma, tanto las que producen directamente las plataformas como las que son el resultado de investigadores independientes con acceso a información privilegiada. Esta categoría agrupa también pronunciamientos abstractos o de *deseos* de las plataformas sobre cómo lidiar con el desafío de la desinformación.

Las empresas no desplegaron todas las acciones a la vez: cada acción tiene su historia, vinculada a la temporalidad del fenómeno de la desinformación y a los cambios de actitud frente a Internet en el eje optimismo / pesimismo. Su objetivo fue enfrentar las demandas de usuarios, de sociedad civil y gobiernos en relación al rol central que adquirieron en el flujo de información en Internet como consecuencia de su propio éxito.

Las acciones de concientización (1) operan en un nivel que implica sólo inversión de recursos de las compañías, pero no impactan de ninguna manera en su modelo de negocios. Lo mismo ocurre, aunque en menor medida, con las acciones de transparencia y relaciones públicas (4), que sólo pueden presentarse como “costosas” desde un punto de vista de la libre competencia (revelan información a los adversarios o competidores). Las acciones de moderación implican el diseño y la implementación de políticas vinculadas a los contenidos, como las *community guidelines* o *terms of service* (TOS), que pueden implicar costos altos dependiendo de la complejidad de la estructura (3). Finalmente, los cambios en el código afectan el uso que los usuarios pueden hacer de los servicios que ofrecen las compañías y, en tanto esas modificaciones pueden tener un efecto disruptivo inesperado, deben verse como las acciones más “radicales” dentro del menú de opciones disponibles (2).²³

²³ El anuncio realizado por Mark Zuckerberg el 19 de noviembre de 2016, unos diez días después de la elección presidencial que llevó a Donald Trump a la presidencia, es bastante revelador de lo que vino luego. Esas medidas fueron las siguientes: (1) detección más robusta, a través de mecanismos automatizados; (2) facilidad en el reporte, para facilitar la “detección”; (3) avisos — Zuckerberg dijo estar “explorando” la posibilidad de etiquetar historias que han sido “reportadas” como falsas por “terceras partes o nuestra comunidad”; (4) alianzas con “fact checkers” y—en general—escuchar más al periodismo y la industria de la verificación; (5) recomendación de artículos “de calidad” (2). Zuckerberg anunció que iba a “levantar el nivel”

1. Acciones de concientización

Son acciones tendientes a alertar a los usuarios sobre las campañas de desinformación o a promover el periodismo de calidad como respuesta presuntamente efectiva al fenómeno. Estas acciones agrupan también las diversas alianzas establecidas por las compañías con otros actores, especialmente con medios tradicionales de comunicación y organizaciones de verificación.

a. Apoyo y promoción del periodismo “de calidad”

Tanto Google como Facebook han apoyado iniciativas tendientes a fortalecer al periodismo “de calidad”. Ambas compañías concentraron esos esfuerzos en dos iniciativas que parecen bastante similares: el *Facebook Journalism Project*²⁴ y la *Google News Initiative*.²⁵

En general, la mayor parte de las acciones de estas iniciativas se verificaron en Estados Unidos, Canadá y el Reino Unido, pero también se han extendido a otras áreas del mundo. En octubre de 2018, por ejemplo, el *Facebook Journalism Project* anunció apoyos a periodistas árabes para luchar contra la desinformación²⁶ y dos años des-

de calidad de los artículos que aparecen en “artículos relacionados”; (6) cambios en las políticas de avisos para “alterar” la “economía de las fake news” (3). Ver M. Zuckerberg (*Mark Zuckerberg On Disinformation*, FACEBOOK STATUS UPDATE, 19/11/2016, disponible en <https://www.facebook.com/zuck/posts/10103269806149061> Fecha de consulta: 2/marzo/2020.).

²⁴ Ver, p.ej., Brown Campbell (*Campbell Brown: Facebook Is Doing More To Support To Local News*, CAMPBELL BROWN: FACEBOOK IS DOING MORE TO SUPPORT TO LOCAL NEWS, 15/01/2019, disponible en <https://www.facebook.com/journalismproject/facebook-supports-local-news> Fecha de consulta: 11/marzo/2020.); J. Mabry; D. Mendoza (*Facebook Develops Training And Support For Local Newsrooms*, FACEBOOK DEVELOPS TRAINING AND SUPPORT FOR LOCAL NEWSROOMS, 27/04/2017, disponible en <https://www.facebook.com/journalismproject/facebook-training-support-local-newsrooms> Fecha de consulta: 2/marzo/2020.); F. Dinkelspiel (*Berkeleyside To Launch News Site In Oakland, Will Convert To Nonprofit*, BERKELEYSIDE, 10/12/2019, disponible en <https://www.berkeleyside.com/2019/12/10/berkeleyside-to-launch-news-site-in-oakland-will-convert-to-non-profit> Fecha de consulta: 27/agosto/2020.). Facebook y Mozilla también apoyaron la *News Integrity Initiative* en la escuela de periodismo de CUNY. Ver Poynter (*Can Trust In The News Be Repaired? Facebook, Craig Newmark, Mozilla And Others Are Spending \$14 Million To Try*, POYNTER, 03/04/2017, disponible en <https://www.poynter.org/tech-tools/2017/can-trust-in-the-news-be-repaired-facebook-craig-newmark-mozilla-and-others-are-spending-14-million-to-try/> Fecha de consulta: 2/marzo/2020.); J. Rutenberg; M. Isaac («Facebook Hires Campbell Brown To Lead News Partnerships Team», THE NEW YORK TIMES, 6/1/2017, disponible en <https://www.nytimes.com/2017/01/06/business/media/facebook-campbell-brown-media-fake-news.html> Fecha de consulta: 2/marzo/2020.).

²⁵ Ver <https://newsinitiative.withgoogle.com/>

²⁶ MENA Herald, *New Program Will Help Arab Journalists Sort Fact From Fiction On Social Media*, MENA HERALD, 28/11/2018, disponible en <https://www.menaherald.com/en/business/media-marketing/new-program-will-help-arab-journalists-sort-fact-fiction-social-media> Fecha de consulta: 11/marzo/2020.

pués anunció la expansión de su programa de “aceleradores” a nivel global.²⁷ En Latinoamérica, parecen haber alcanzado a algunos medios de Brasil.²⁸ La *Google News Initiative* estuvo detrás del lanzamiento de *Reverso* en la Argentina²⁹ y apoyó a *Impacto.jor*, un proyecto dirigido a medir el impacto de los medios de comunicación.

Este tipo de acciones se sostienen sobre una premisa: los medios tradicionales de comunicación pueden ser actores importantes en luchar contra la desinformación. Sin embargo, ello no resulta para nada obvio. Por ejemplo, sería necesario analizar contextos particulares para comprender si esos medios retienen su rol de *gatekeepers* y en qué medida. Por otro lado, tampoco es evidente que los medios de comunicación tradicionales operen siempre como contraposición a las campañas de desinformación: por ejemplo, el estudio de Benkler, Faris y Roberts sobre el flujo de información en Estados Unidos reveló que una cadena tradicional de noticias como *Fox News* cumplió un rol central en apalancar la estrategia de desinformación del presidente Trump frente a la investigación de Robert Müller.³⁰

b. Concientizar usuarios

Otras acciones de las plataformas buscaron concientizar a los usuarios sobre la existencia de la *desinformación* misma. Este tipo de acciones descansa sobre el paradigma señalado en la sección anterior que proponía que le toca a los ciudadanos definir qué es verdadero y qué es falso. El objetivo parece doble: alfabetizar a usuarios desprevenidos o sorprendidos por el caótico flujo de información que parece acelerarse día a día; y recordar que los ciudadanos de una comunidad política democrática tienen el deber de abordar críticamente el debate público.

Así, en abril de 2017 Facebook, en colaboración con First Draft, lanzó en cator-

²⁷ D. Grant, *What's Next For Facebook's Accelerator Program*, FACEBOOK ACCELERATOR PROGRAM: WHAT'S PLANNED FOR 2020, 28/01/2020, disponible en <https://www.facebook.com/journalismproject/programs/accelerator/whats-next-2020> Fecha de consulta: 13/marzo/2020.

²⁸ Cf. *Ibid.*

²⁹ P. J. Blanco, «La Elección De Las “fake News”: Más De 100 Medios Lanza “reverso”, Para Combatir La Desinformación En Campaña», DIARIO CLARIN, 11/6/2019, disponible en https://www.clarin.com/politica/eleccion-fake-news-100-medios-lanza-reverso-grupo-combatir-desinformacion-campana_0_QYwSO8kRP.html Fecha de consulta: 13/marzo/2020.

³⁰ Y. BENKLER; R. FARIS; H. ROBERTS, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, New York, NY, 2018, pág. 158 («...vemos a Fox News tomando un rol central de una forma clara y distintiva, desviando la atención y la culpa, interpretando la investigación en términos profundamente partidarios, y sembrando confusión dudas. Y es aquí también dónde el marco amplio y ambiguo adquiere caracteres de falsificación, cuyos patrones encajan con los patrones de campañas sostenidas de desinformación por actores de propaganda, más que con errores episódicos de un actor periodístico...»).

ce países una serie de “consejos” para identificar noticias falsas.³¹ En el mismo mes, lanzó campañas en diarios alemanes para concientizar sobre el fenómeno.³² YouTube también anunció la realización de talleres de trabajo con adolescentes en el Reino Unido³³ y Google lanzó una iniciativa similar a nivel global.³⁴ Twitter también promovió una *Media Literacy Week* en algunos de sus mercados.³⁵

Este tipo de acciones son poco problemáticas pero la evidencia es contradictoria en cuanto a su efectividad. Se ha demostrado que, en ocasiones, los esfuerzos por contrarrestar información falsa pueden ser contraproducentes.³⁶ Sin embargo, la intervención preventiva parece tener efectos positivos en algunos casos, ya que pone a las personas a la defensiva desde el punto de vista cognitivo y podría ser efectivo en impedir posibles efectos epistémicos de las campañas de desinformación.³⁷

c. Alianzas con verificadores y otros actores

³¹ A. Mosseri, *A New Educational Tool Against Misinformation*, ABOUT FACEBOOK, 06/04/2017, disponible en <https://about.fb.com/news/2017/04/a-new-educational-tool-against-misinformation/> Fecha de consulta: 2/marzo/2020 («Improving news literacy is a global priority, and we need to do our part to help people understand how to make decisions about which sources to trust»); C. Silverman, *Facebook Wants To Teach You How To Spot Fake News On Facebook*, BUZZFEED NEWS, 06/04/2017, disponible en <https://www.buzzfeednews.com/article/craigsilverman/facebook-wants-to-teach-you-how-to-spot-fake-news-on#.erZ2mjEY2a> Fecha de consulta: 2/marzo/2020.

³² «Facebook Buys Full-page Ads In Germany In Battle With Fake News», BLOOMBERG.COM, 13/4/2017, disponible en <https://www.bloomberg.com/news/articles/2017-04-13/facebook-buys-full-page-ads-in-germany-in-battle-with-fake-news> Fecha de consulta: 2/marzo/2020.

³³ BBC, *Youtube To Offer Fake News Workshops To Teenagers*, BBC NEWSBEAT, 20/04/2017, disponible en <http://www.bbc.co.uk/newsbeat/article/39653429/youtube-to-offer-fake-news-workshops-to-teenagers> Fecha de consulta: 2/marzo/2020 («Naomi Gummer, head of public policy at YouTube UK, said: “The internet is what we want it to be. It can be an unpleasant place where people misunderstand and deliberately deceive each other. Or it can be this amazing place where we can share, collaborate, understand and help each other”»).

³⁴ P. Diwanji, «be Internet Awesome»: Helping Kids Make Smart Decisions Online, GOOGLE, 06/06/2017, disponible en <https://blog.google/technology/families/be-internet-awesome-helping-kids-make-smart-decisions-online/> Fecha de consulta: 4/marzo/2020.

³⁵ Twitter public policy (*Update: Russian Interference In The 2016 Us Presidential Election*, TWITTER BLOG, 28/09/2017, disponible en https://blog.twitter.com/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html Fecha de consulta: 4/marzo/2020.); ver <https://medialiteracyweek.us/wp-content/uploads/2015/07/2017-media-literacy-week-press-release-final.pdf>.

³⁶ D. J. Flynn; B. Nyhan; J. Reifler, «The Nature And Origins Of Misperceptions: Understanding False And Unsupported Beliefs About Politics», *POLITICAL PSYCHOLOGY*, vol. 38, S1, 2017, Disponible en <https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12394>; Ver B. Nyhan; J. Reifler, «Estimating Fact-checking's Effects». American Press Institute, Arlington, V.A. 2015.

³⁷ Cf. J. COOK; S. LEWANDOWSKY, «Misinformation And How To Correct It», en Robert A Scott, Stephan M Kosslyn (eds.) *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1, Wiley, 2015, págs. 6-7 Disponible en <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118900772>. Además de esta estrategia, Cook y Lewandowski señalan que la fuerza con la cual se refuta información falsa y la explicación alternativa que «llena el vacío» que deja la información retirada de los modelos mentales que se utilizan para comprender la información de que es trate fueron efectivos, en condiciones experimentales, para fortalecer el efecto de la refutación.

Las alianzas con organizaciones verificadoras de información parecen haberse instalado como la respuesta preferida por las plataformas para afrontar el fenómeno de la desinformación. Facebook anunció alianzas con la *Associated Press*,³⁸ con *Boom* para trabajar en el mercado de la India,³⁹ con el IFCN para fortalecer el chequeo de noticias⁴⁰ y con la agencia Reuters.⁴¹ En octubre de 2018 Google lanzó un buscador especializado en verificadores, e incluye a actores latinoamericanos.⁴² La colaboración entre las plataformas y las agencias verificadoras, que proliferaron en los últimos tiempos, parece ser estrecha, como lo demuestra el proyecto *Reverso* en la Argentina.

Este tipo de alianzas tampoco presentan problemas significativos desde el punto de vista de la libertad de expresión, y se han reportado efectos positivos.⁴³ Sin embargo, subsisten numerosos desafíos, entre ellos, su limitado impacto en las personas menos informadas o con menor educación, quienes—a la vez—son más proclives a consumir información falsa.⁴⁴

2. Acciones de código

En este grupo se identifican las acciones que modifican de alguna manera el *código* de la plataforma. La detección, reporte y recomendaciones de información adicional son acciones típicas de esta categoría.

d. Detección de información falsa

El modelo de producción horizontal de información transforma a las grandes plataformas en *canalizadoras* de contenidos creados por terceros. . Pero, como

³⁸ AP, *Ap To Debunk Election Misinformation On Facebook*, ASSOCIATED PRESS, 07/03/2018, disponible en <https://www.ap.org/press-releases/2018/ap-to-debunk-election-misinformation-on-facebook> Fecha de consulta: 9/marzo/2020.

³⁹ P. Dixit, *Facebook Has Finally Started Fact-checking Fake News In Its Largest Market*, BUZZFEED NEWS, 17/04/2018, disponible en <https://www.buzzfeednews.com/article/pranavdixit/facebook-has-finally-started-fact-checking-fake-news-in-its> Fecha de consulta: 9/marzo/2020.

⁴⁰ Poynter, *Ifcn And The Facebook Journalism Project Announce Fact-checking Innovation Initiative*, POYNTER, 05/11/2019, disponible en <https://www.poynter.org/fact-checking/2019/innovation-initiative/> Fecha de consulta: 27/agosto/2020.

⁴¹ D. Patadia, «Reuters Launches Fact-checking Initiative To Identify Misinformation, In Partnership With Facebook», REUTERS, 12/2/2020, disponible en <https://www.reuters.com/article/rpb-fbfactchecking-idUSKBN2061TG> Fecha de consulta: 27/agosto/2020.

⁴² Disponible en: <https://toolbox.google.com/factcheck/>

⁴³ B. Nyhan; J. Reifler, «Estimating Fact-checking's Effects», cit.

⁴⁴ Cf. G. Pennycook; D. G. Rand, «Who Falls For Fake News? The Roles Of Analytic Thinking, Motivated Reasoning, Political Ideology, And Bullshit Receptivity», SSRN ELECTRONIC JOURNAL, 2017, Disponible en <https://www.ssrn.com/abstract=3023545>.

se señaló antes, de manera creciente ejercen un rol de curación y moderación activa (Tabla 1). Las empresas le han dado forma a sus políticas de moderación como consecuencia de sus propios intereses económicos y por presiones de actores externos: las exigencias legales en diversos países para combatir discursos claramente prohibidos como la pornografía infantil o las amenazas las presiones sociales por combatir discursos problemáticos pero no necesariamente ilegales como el llamado discurso discriminatorio las presiones políticas para—por ejemplo—abordar fenómenos como el que éste estudio analiza.

Tabla 1 - Razones que empujan a la moderación de contenidos por parte de intermediarios.

Razones	Positivas	Negativas
Empresariales	la introducción de los algoritmos como forma de generar contenido más relevante y aumentar el engagement de usuarios	cuando las empresas prefieren que ciertos discursos no estén en sus plataformas
Sociales	cuando la sociedad pide por ciertos discursos (p.ej. campañas de concientización de bien público)	cuando la sociedad rechaza ciertos discursos y pide acciones contra ellos (p.ej. d. de odio)
Políticas	las que se originan en el diálogo y las alianzas (p.ej. autoridades electorales)	las que derivan de presiones del sistema político (investigaciones legislativas)
Legales	las demandas que exigen ciertos contenidos o prácticas	las que establecen obligaciones en términos de discursos prohibidos

Es en el marco de estas acciones generales de moderación que deben ubicarse las acciones tendientes a “detectar” información falsa. Son acciones que recurren a diversos métodos, pero parten de la misma premisa: la desinformación es un tipo de discurso indeseable en estos servicios.⁴⁵ En la implementación de esa política, el primer desafío siempre fue identificar con algún tipo de precisión el contenido que es falso y para ello las plataformas parecen haber recurrido a dos

⁴⁵ Este evento parece estar vinculado a la elección presidencial de 2016 en los Estados Unidos de 2016. Ver T. GILLESPIE, *Custodians Of The Internet: Platforms, Content Moderation, And The Hidden Decisions That Shape Social Media*, Yale University Press, New Haven, 2018, págs. 65-66. El punto es interesante porque existió cierta ambigüedad de parte de las plataformas respecto de sus propias capacidades para «monitorear» esos contenidos.

estrategias: el reporte de los usuarios y la detección automática.

La primera aproximación simplemente pide a los usuarios que reporten contenido que es sospechado como falso, p.ej., a través de “encuestas” en las que se solicita ese *feedback*⁴⁶ o estableciendo la posibilidad de “reportar” contenido como falso. Google implementó la posibilidad de reportar las recomendaciones automáticas al introducir texto en el buscador⁴⁷ y Twitter estableció—en abril de 2019—una nueva modalidad de reporte que incluye “información falsa respecto de elecciones”,⁴⁸ aunque esta opción aún no se encuentra disponible en América Latina.⁴⁹

La segunda aproximación se basa en el desarrollo de herramientas tecnológicas capaces de detectar el contenido falso de manera automática, con base en algoritmos, inteligencia artificial y *machine learning*. Así, por ejemplo, Facebook anunció la adopción de criterios basados en los comportamientos de los usuarios para catalogar contenidos como sensacionalistas o no sensacionalistas⁵⁰ (aunque cambió su aproximación después de que una publicación satírica fuera incorrectamente catalogada como falsa⁵¹). Por otro lado, YouTube en enero de 2019 anunció que buscaría “bajarle el volumen” a contenidos que si bien no violan sus políticas internas, se encuentran “cerca” de violarlas.⁵² La apuesta por este tipo

⁴⁶ Vr C. Zang; S. Chen, *Using Qualitative Feedback To Show Relevant Stories*, ABOUT FACEBOOK, 01/02/2016, disponible en <https://about.fb.com/news/2016/02/news-feed-fyi-using-qualitative-feedback-to-show-relevant-stories/> Fecha de consulta: 2/marzo/2020.

⁴⁷ Google's «project Owl» -- A Three-pronged Attack On Fake News & Problematic Content, SEARCH ENGINE LAND, 25/04/2017, disponible en <https://searchengineland.com/googles-project-owl-attack-fake-news-273700> Fecha de consulta: 7/mayo/2020.

⁴⁸ Twitter Safety, *Strengthening Our Approach To Deliberate Attempts To Mislead Voters*, TWITTER BLOG, 24/04/2019, disponible en https://blog.twitter.com/en_us/topics/company/2019/strengthening-our-approach-to-deliberate-attempts-to-mislead-vot.html Fecha de consulta: 11/marzo/2020.

⁴⁹ Hugo Rodríguez Nicolat, ‘Comentarios de Twitter a La Encuesta de Investigación: “Desinformación En Internet: Acciones y Reacciones”’ (2020, en archivo con los autores), págs.4-5 (“Es importante que esta política sea aplicada de manera global y si bien la herramienta como tal no está disponible todavía en Latinoamérica, el plan es implementarla en todo el mundo tan pronto nos sea posible. Debemos mencionar no obstante que la ausencia de la herramienta de reporte público hasta el momento, esto en sentido alguno implica la ausencia de aplicación de esta política, en tanto: i) Twitter aplica sus reglas de manera proactiva (más del 50% de los contenidos accionados por nosotros son detectados de manera proactiva, sin mediar reporte de algún usuario) y ii) el equipo de Política Pública para Latinoamérica Hispanoparlante trabaja constantemente para establecer un canal de comunicación directo con las diferentes autoridades electorales de la región (por ejemplo Argentina, Colombia y México) para reportar contenido violatorio de nuestras políticas de contenido publicitario o de las regulaciones locales en este sentido”).

⁵⁰ M. Zuckerberg, *Building Global Community*, FACEBOOK, 16/02/2017, disponible en <https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634> Fecha de consulta: 2/marzo/2020.

⁵¹ E. Wemple, «Facebook Working On Approach To Classifying Satirical News Pieces - The Washington Post», THE WASHINGTON POST, 5/3/2018, disponible en <https://www.washingtonpost.com/blogs/erik-wemple/wp/2018/03/05/facebook-working-on-approach-to-classifying-satirical-news-pieces/> Fecha de consulta: 7/mayo/2020.

⁵² YouTube, *Continuing Our Work To Improve Recommendations On Youtube*, OFFICIAL YOUTUBE BLOG, 25/01/2019, disponible en <https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html> Fecha de consulta: 11/marzo/2020.

de herramientas está basada en la potencial escalabilidad de una solución exitosa. Sin embargo es difícil: detectar contenido falso es complejo e incluso cuando es relativamente fácil (por ejemplo, en el caso de los *deepfakes*) es muy probable que requiera de la intervención humana para, entre otras cosas, distinguir una campaña de desinformación de contenido meramente satírico.⁵³

e. Contextualización de información mediante el “etiquetado” de contenidos

Otra acción que diversas plataformas han adoptado en los últimos tiempos tiene que ver con el “etiquetado” de contenidos para dar más contexto a la información. Así, por ejemplo, en junio de 2019 Twitter anunció que etiquetaría publicaciones de políticos o figuras públicas que violan sus reglas y que, en otras circunstancias, serían dados de baja.⁵⁴ Esta acción se volvió especialmente polémica cuando empezó a etiquetar publicaciones del entonces presidente de los Estados Unidos, Donald Trump.⁵⁵ Facebook anunció en junio de 2020 que etiquetaría a medios que pertenecen a estados extranjeros como p.ej., la agencia *Sputnik* de Rusia⁵⁶ y en agosto Twitter hizo un anuncio similar.⁵⁷ Asimismo, las plataformas también anunciaron que identificarían los lugares de origen de diversos conteni-

⁵³ Ver, p.ej., un reporte de avance de Facebook en este sentido en T. Lyons (*Increasing Our Efforts To Fight False News*, ABOUT FACEBOOK, 21/06/2018, disponible en <https://about.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/> Fecha de consulta: 14/julio/2020.). Sobre los desafíos de la IA para luchar contra la desinformación, ver S. C. Woolley (*We're Fighting Fake News Ai Bots By Using More Ai. That's A Mistake.*, MIT TECHNOLOGY REVIEW, 08/01/2020, disponible en <https://www.technologyreview.com/2020/01/08/130983/were-fighting-fake-news-ai-bots-by-using-more-ai-thats-a-mistake/> Fecha de consulta: 14/julio/2020.)

⁵⁴ E. Dwoskin; T. Romm «*Twitter Adds Labels For Tweets That Break Its Rules — A Move With Potentially Stark Implications For Trump's Account*», *The Washington Post*, 27/06/2019, disponible en <https://www.washingtonpost.com/technology/2019/06/27/twitter-adds-labels-tweets-that-break-its-rules-putting-president-trump-companys-crosshairs/> Fecha de consulta: 11/marzo/2020.

⁵⁵ D. Alba; K. Conger; R. Zhong, «*Twitter Adds Warnings To Trump And White House Tweets, Fueling Tensions*», *THE NEW YORK TIMES*, 29/5/2020, disponible en <https://www.nytimes.com/2020/05/29/technology/trump-twitter-minneapolis-george-floyd.html> Fecha de consulta: 14/julio/2020; J. Byrnes, *Twitter Flags Trump Tweet On Protesters For Including «threat Of Harm»*, *THEHILL*, 23/06/2020, disponible en <https://thehill.com/policy/technology/504133-twitter-flags-trump-tweet-on-protesters-for-including-threat-of-harm> Fecha de consulta: 14/julio/2020; K. Conger, «*Twitter Had Been Drawing A Line For Months When Trump Crossed It*», *THE NEW YORK TIMES*, 30/5/2020, disponible en <https://www.nytimes.com/2020/05/30/technology/twitter-trump-dorsey.html> Fecha de consulta: 14/julio/2020; A. Wiener, *Trump, Twitter, Facebook, And The Future Of Online Speech*, *THE NEW YORKER*, 06/07/2020, disponible en <https://www.newyorker.com/news/letter-from-silicon-valley/trump-twitter-facebook-and-the-future-of-online-speech> Fecha de consulta: 10/julio/2020.

⁵⁶ N. Gleicher, *Labeling State-controlled Media On Facebook*, ABOUT FACEBOOK, 04/06/2020, disponible en <https://about.fb.com/news/2020/06/labeling-state-controlled-media/> Fecha de consulta: 27/agosto/2020.

⁵⁷ Twitter Support, *New Labels For Government And State-affiliated Media Accounts*, TWITTER BLOG, 06/08/2020, disponible en https://blog.twitter.com/en_us/topics/product/2020/new-labels-for-government-and-state-affiliated-media-accounts.html Fecha de consulta: 27/agosto/2020.

dos.⁵⁸ Twitter también anunció que en ciertos casos pediría a usuarios que ratifiquen su intención de re-publicar cierta información antes de subirla.⁵⁹

La “contextualización” de información llegó a niveles nunca vistos durante las elecciones presidenciales en Estados Unidos de 2020 cuando comenzaron a etiquetar publicaciones del ex presidente Donald J. Trump que deliberadamente desinformaban sobre un supuesto “fraude” en las elecciones.⁶⁰

Este último caso pone de presente algunos de los problemas involucrados en estas acciones de contextualización. Si bien en sí misma si se limita a dar más contexto es poco riesgosa desde el punto de vista de la libertad de expresión, podría ser problemática si ese etiquetado tuviera efectos en la forma en que circula la información; por ejemplo, si el algoritmo de recomendación de contenidos considerara negativamente a esas etiquetas. Por otro lado, estas acciones de moderación enfrentan un desafío considerable en términos de escala, replicabilidad y *gestión* de estándares utilizados para señalar a unos y otros: la aplicación homogénea de la norma en distintos países y en distintos contextos promete ser un desafío considerable y numerosos críticos han señalado que mientras los *tuits* de Trump merecían el etiquetado de la compañía, *tuits* similares de otros líderes políticos o de opinión incluso no recibían el mismo trato.

f. Más periodismo de calidad y verificación

A veces las alianzas con verificadores suponen cambios en el código de las plataformas. Google anunció, por ejemplo, que los verificadores recibirían un lugar

⁵⁸ A. Joseph, *Making Pages And Accounts More Transparent*, ABOUT FACEBOOK, 22/04/2020, disponible en <https://about.fb.com/news/2020/04/page-and-account-transparency/> Fecha de consulta: 27/agosto/2020; R. Wong, *Google's Taking Another Big Step To Stop The Spread Of Fake News*, MASHABLE, 17/12/2017, disponible en <https://mashable.com/2017/12/17/google-news-no-hiding-country-origin-stop-fake-news/> Fecha de consulta: 6/marzo/2020.

⁵⁹ Twitter Support, *Twitter Support On Twitter: "Sharing An Article Can Spark Conversation, So You May Want To Read It Before You Tweet It. To Help Promote Informed Discussion, We're Testing A New Prompt On Android — When You Retweet An Article That You Haven't Opened On Twitter, We May Ask If You'd Like To Open It First."*, TWITTER, 10/06/2020, disponible en <https://twitter.com/TwitterSupport/status/1270783537667551233> Fecha de consulta: 27/agosto/2020.

⁶⁰ La situación escaló de manera persistente, manteniendo el etiquetado como herramienta de preferencia, hasta que, el 6 de enero de 2021, una manifestación ante el Congreso de los Estados Unidos avanzó sobre las fuerzas de seguridad e ingresó al recinto donde se estaban validando los resultados de las elecciones. A partir del incidente diversas plataformas suspendieron la cuenta del presidente por considerar que había incitado a esa “insurrección” en parte a través de su discurso desinformador. En el caso de Facebook, su Oversight Board decidió tomar el caso para expedirse al respecto. Ver S. Tavernise; M. Rosenberg, «These Are the Rioters Who Stormed the Nation's Capitol», *The New York Times*, 8/1/2021, disponible en <https://www.nytimes.com/2021/01/07/us/rioters-capitol.html> Fecha de consulta: 2/febrero/2021; D. Ghosh; J. Hendrix, *Facebook's Oversight Board Takes on the Curious Case of Donald J. Trump*, *Verfassungsblog*, 29/01/2021, disponible en <https://verfassungsblog.de/fob-trump/> Fecha de consulta: 2/febrero/2021.

privilegiado en los resultados de la búsqueda cuando la información referida haya sido debidamente chequeada por un verificador reconocido,⁶¹ un cambio que—más adelante—introdujo también en Google News.⁶² También anunció que daría más lugar al periodismo de calidad en su buscador,⁶³ algo que también anunció Facebook en relación a cambios en su algoritmo.⁶⁴

En abril de 2017 Facebook anunció que los “artículos relacionados” tendrían un lugar más preeminente, antes de que los usuarios lean determinados contenidos.⁶⁵ Ello incluiría la promoción proactiva de información verificada cuando los mecanismos automatizados de detección enviasen a verificadores independientes contenidos sospechosos y—luego del correspondiente “chequeo”—podrían mostrar esos artículos debajo del artículo original.⁶⁶ Asimismo, adoptó acciones en contra del *cloaking*, una técnica para evitar ciertos controles internos,⁶⁷ el *spamming*⁶⁸ y el *clickbait*.⁶⁹ En octubre de 2017 Facebook también anunció un nuevo botón, en etapa de prueba, para ofrecer más información de contexto en artículos compar-

⁶¹ Cf. A. Mantzaris (*Google Is Now Highlighting Fact Checks In Search*, POYNTER, 07/04/2017, disponible en <https://www.poynter.org/fact-checking/2017/google-is-now-highlighting-fact-checks-in-search/> Fecha de consulta: 2/marzo/2020.); A. Mantzaris (*Google News Now Has A «fact Check» Tag*, POYNTER, 13/10/2016, disponible en <https://www.poynter.org/fact-checking/2016/google-news-now-has-a-fact-check-tag/> Fecha de consulta: 2/marzo/2020.). El comportamiento no pudo verificarse, depende del uso de diversas tecnologías como p.ej. schema.org

⁶² J. Lichterman, *Google News Launches A Streamlined Redesign That Gives More Prominence To Fact Checking*, NIEMAN LAB, 27/06/2017, disponible en <https://www.niemanlab.org/2017/06/google-news-launches-a-streamlined-redesign-that-gives-more-prominence-to-fact-checking/> Fecha de consulta: 4/marzo/2020.

⁶³ R. Gingras, *Elevating Original Reporting In Search*, GOOGLE BLOG, 12/09/2019, disponible en <https://blog.google/products/search/original-reporting> Fecha de consulta: 11/marzo/2020; D. Osborn, *Smarter Organization Of Top Stories In Search*, GOOGLE, 11/12/2019, disponible en <https://blog.google/products/search/smarter-organization-top-stories-search/> Fecha de consulta: 27/agosto/2020.

⁶⁴ S. Fischer, *Axios Media Trends: Facebook To Change Its Algorithm*, AXIOS, 30/06/2020, disponible en <https://www.axios.com/newsletters/axios-media-trends-40d2f971-c434-4359-8a0f-7fa4155626bc.html> Fecha de consulta: 27/agosto/2020.

⁶⁵ S. Su, *New Test With Related Articles*, ABOUT FACEBOOK, 25/04/2017, disponible en <https://about.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/> Fecha de consulta: 2/marzo/2020.

⁶⁶ *Ibid.* («Now, we will start using updated machine learning to detect more potential hoaxes to send to third-party fact checkers. If an article has been reviewed by fact checkers, we may show the fact checking stories below the original post. In addition to seeing which stories are disputed by third-party fact checkers, people want more context to make informed decisions about what they read and share. We will continue testing updates to Related Articles and other ongoing News Feed efforts to show less false news on Facebook and provide people context if they see false news»).

⁶⁷ R. Leathern, *Addressing Cloaking So People See More Authentic Posts*, ABOUT FACEBOOK, 09/08/2017, disponible en <https://about.fb.com/news/2017/08/news-feed-fyi-addressing-cloaking-so-people-see-more-authentic-posts/> Fecha de consulta: 4/marzo/2020.

⁶⁸ A. Mosseri, *Showing More Informative Links In News Feed*, ABOUT FACEBOOK, 30/06/2017, disponible en <https://about.fb.com/news/2017/06/news-feed-fyi-showing-more-informative-links-in-news-feed/> Fecha de consulta: 4/marzo/2020.

⁶⁹ A. Babu; A. Liu; J. Zhang, *New Updates To Reduce Clickbait Headlines*, ABOUT FACEBOOK, 17/05/2017, disponible en <https://about.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/> Fecha de consulta: 4/marzo/2020.

tidos,⁷⁰ algo que luego se convirtió en el *Trust Indicators*.⁷¹ Google también desarrolló una herramienta similar, (*Knowledge Panels*) que permite en su buscador conocer más respecto del origen de la información.⁷² En marzo de 2018, YouTube anunció que vincularía información de Wikipedia a videos que promovieran teorías conspirativas⁷³ y en febrero de 2019 hizo un anuncio similar respecto de la India, vinculado a la publicación de información de verificadores.⁷⁴

g. Prohibición de comportamientos

Una de las estrategias preferidas por las plataformas para combatir la desinformación es apuntar a “comportamientos” de usuarios que puedan estar relacionados con campañas de desinformación. Así, por ejemplo, es posible que ese tipo de campañas estén motorizadas por cuentas anónimas, que dejen rastros de comportamientos específicos—una tendencia irrazonable a *retuitear*, a enviar o esparcir información, a buscar instalar tendencias—, o que simplemente se trate de *bots*, es decir, cuentas que actúan a partir de determinado código, de manera automática (por ejemplo, *retuiteando* ciertas cuentas o contenidos o publicando información periódica de determinadas características, etcétera). Esta estrategia es en cierta medida “preferida” por las plataformas porque permiten evitar el espinoso problema de la definición de la desinformación: ya no se vuelve necesario determinar si cierto contenido es falso o no, sino que basta con detectar comportamientos prohibidos, una tarea que puede realizarse sobre bases objetivas y mediante procesos automatizados.

Twitter señaló este punto especialmente.⁷⁵ Dentro de sus estrategias principales para combatir la desinformación se incluyen no sólo la detección de comporta-

⁷⁰ A. Anker, *New Test To Provide Context About Articles*, ABOUT FACEBOOK, 05/10/2017, disponible en <https://about.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/> Fecha de consulta: 6/marzo/2020.

⁷¹ A. Anker, *Launching New Trust Indicators From The Trust Project For News On Facebook*, FACEBOOK FOR MEDIA, 17/11/2017, disponible en <https://www.facebook.com/facebookmedia/blog/launching-new-trust-indicators-from-the-trust-project-for-news-on-facebook> Fecha de consulta: 6/marzo/2020 («We are initially testing these Trust Indicators with a small group of publishers, with plans to expand more broadly over the coming months»).

⁷² Google, *Learn More About Publishers On Google*, GOOGLE, 17/11/2017, disponible en <https://blog.google/products/search/learn-more-about-publishers-google/> Fecha de consulta: 13/marzo/2020.

⁷³ C. Newton, *Youtube Will Add Information From Wikipedia To Videos About Conspiracies*, THE VERGE, 13/03/2018, disponible en <https://www.theverge.com/2018/3/13/17117344/youtube-information-cues-conspiracy-theories-susan-wojcicki-sxsw> Fecha de consulta: 9/marzo/2020.

⁷⁴ Dixit, Pranav, *Youtube Is Rolling Out A Feature That Shows Fact Checks When People Search For Sensitive Topics*, BUZZFEED NEWS, 07/03/2019, disponible en <https://www.buzzfeednews.com/article/pranavdixit/youtube-debunk-information-panels-india> Fecha de consulta: 11/marzo/2020.

⁷⁵ Twitter public policy, *Update*, cit.

mientos que revelen prácticas que violan los TOS de la compañía, sino también intentos por impedir que actores malintencionados le “ganen” al sistema p.ej., manipulando los *trending topics*.⁷⁶ Al respecto, Twitter informó:

“Como manipulación de la plataforma entendemos utilizar Twitter con el propósito de amplificar o suprimir información de forma artificial o llevar a cabo acciones que manipulen u obstaculicen la experiencia de los usuarios en Twitter. En este sentido, prohibimos llevar a cabo acciones masivas, de alta intensidad o engañosas que confunden a los usuarios o que obstaculizan su experiencia. Hay muchas formas de manipular la plataforma, y nuestras reglas están destinadas a contrarrestar una gran variedad de comportamientos prohibidos, por ejemplo: spam con fines comerciales, que generalmente apunta a desviar el tráfico o la atención de una conversación en Twitter hacia otras cuentas, sitios web, productos, servicios o iniciativas; interacciones falsas que apuntan a hacer que las cuentas o los contenido parezcan más populares o activos de lo que realmente son; y actividades coordinadas que apuntan a influir de forma artificial las conversaciones mediante el uso de varias cuentas, cuentas falsas, acciones automáticas o secuencias de comandos”.⁷⁷

Facebook opta por un modelo similar, pero basado en una política diferente: allí las cuentas anónimas están prohibidas — Facebook desea que cada uno de sus usuarios ate a su cuenta su identidad real.⁷⁸ Al igual que en el caso de Twitter, las medidas adoptadas por Facebook en este sentido le permite evitar tener que analizar el contenido de las publicaciones, lo que lleva a (a) una herramienta automatizada de moderación que (b) no depende de complicados análisis de conte-

⁷⁶ N. Confessore; G. J. X. Dance, «Battling Fake Accounts, Twitter To Slash Millions Of Followers», THE NEW YORK TIMES, 11/7/2018, disponible en <https://www.nytimes.com/2018/07/11/technology/twitter-fake-followers.html> Fecha de consulta: 11/marzo/2020; Cf. Twitter public policy, *Update*, cit.

⁷⁷ Hugo Rodríguez Nicolat, ‘Comentarios de Twitter a La Encuesta de Investigación: “Desinformación En Internet: Acciones y Reacciones” (2020, en archivo con los autores), cit., pág. 2.

⁷⁸ S. Shaik, *Improvements In Protecting The Integrity Of Activity On Facebook*, FACEBOOK, 13/04/2017, disponible en <https://www.facebook.com/notes/facebook-security/improvements-in-protecting-the-integrity-of-activity-on-facebook/10154323366590766> Fecha de consulta: 2/marzo/2020 («People come to Facebook to make meaningful connections. From the beginning, we’ve believed that can only be possible if the interactions here are authentic – and if people use the names they’re known by. We’ve found that when people represent themselves on Facebook the same way they do in real life, they act responsibly. Fake accounts don’t follow this pattern, and are closely related to the creation and spread of spam. That’s why we’re so focused on keeping these inauthentic accounts and their activity off our platform»).

nido.⁷⁹ Así, por ejemplo, en agosto de 2017 la compañía anunció que bloquearía anuncios de páginas que de manera “repetida” comparten información falsa.⁸⁰

Este tipo de estrategias parecen haber sido adoptadas para lidiar con los “falsos amplificadores”, los esfuerzos por viralizar cierta información mediante técnicas diversas, como la creación de cuentas falsas, la coordinación de ciertos mensajes, *memes*, etcétera.⁸¹ En el caso de Facebook—que prohíbe las cuentas anónimas—la remoción de cuentas falsas se convierte en un gran *proxy* para afrontar este tipo de comportamientos, que en redes sociales como Twitter (que protege el derecho de los usuarios al anonimato) parece ser más complicado.⁸² Asimismo, Facebook también prohibió “links” en avisos e historias falsas.⁸³

En Twitter, la lucha contra los *bots* se justifica cuando los mismos buscan “socavar el núcleo de las funcionalidades del servicio”.⁸⁴ En noviembre de 2017, Google anunció que dejaría de permitir que sitios que no revelan su origen aparezcan en Google News.⁸⁵ En marzo de 2018, anunció una serie de medidas de restricción sobre anunciantes que abusaban de las políticas de las compañías.⁸⁶

⁷⁹ *Ibid.* («We've made improvements to recognize these inauthentic accounts more easily by identifying patterns of activity — without assessing the content itself. For example, our systems may detect repeated posting of the same content, or an increase in messages sent. With these changes, we expect we will also reduce the spread of material generated through inauthentic activity, including spam, misinformation, or other deceptive content that is often shared by creators of fake accounts»); J. Weedon; W. Nuland; A. Stamos, «Information Operations And Facebook». Facebook, Palo Alto, California. 27 de abril de 2017. Págs. 10 («Through technical advances, we are increasing our protections against manually created fake accounts and using new analytical techniques, including machine learning, to uncover and disrupt more types of abuse. We build and update technical systems every day to make it easier to respond to reports of abuse, detect and remove spam, identify and eliminate fake accounts, and prevent accounts from being compromised. We've made recent improvements to recognize these inauthentic accounts more easily by identifying patterns of activity — without assessing account contents themselves.10 For example, our systems may detect repeated posting of the same content, or aberrations in the volume of content creation. In France, for example, as of April 13, these improvements recently enabled us to take action against over 30, 000 fake accounts”).

⁸⁰ S. Shukla; T. Lyons, *Blocking Ads From Pages That Repeatedly Share False News*, ABOUT FACEBOOK, 28/08/2017, disponible en <https://about.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/> Fecha de consulta: 4/marzo/2020 («Today's update helps to disrupt the economic incentives and curb the spread of false news, which is another step towards building a more informed community on Facebook»).

⁸¹ Cf. J. Weedon; W. Nuland; A. Stamos, «Information Operations And Facebook», cit., pág. 9.

⁸² Cf. *Ibid.* («In the case of Facebook, we have observed that most false amplification in the context of information operations is not driven by automated processes, but by coordinated people who are dedicated to operating inauthentic account»).

⁸³ S. Shukla; T. Lyons, «Blocking Ads From Pages That Repeatedly Share False News», cit.

⁸⁴ C. Crowell, *Our Approach To Bots And Misinformation*, TWITTER BLOG, 14/06/2017, disponible en https://blog.twitter.com/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html Fecha de consulta: 4/marzo/2020.

⁸⁵ R. Wong, *Google's Taking Another Big Step To Stop The Spread Of Fake News*, cit.

⁸⁶ I. Lunden (*Google Removed 2.3b Bad Ads, Banned Ads On 1.5m Apps + 28m Pages, Plans New Policy Manager This Year*, TECHCRUNCH, 14/03/2018, disponible en <http://social.techcrunch.com/2019/03/13/google-removed-2-3b-bad-ads-1-5m-bad-apps-and-28m-bad-pages-plans-new-policy-manager-this-year/> Fecha de consulta: 9/marzo/2020.); S. Spencer (*An Advertising Ecosystem That Works For Everyone*, GOOGLE - THE KEYWORD, 14/03/2018, disponible en <https://blog.google/technology/ads/advertising-ecosystem-works-everyone/> Fecha de consulta: 9/marzo/2020.)

En la medida en que las campañas de desinformación se montan sobre determinadas estrategias de diseminación de información, es posible “combatir” a la desinformación limitando o luchando contra esas estrategias. Pero ello implica dos problemas. En primer lugar, la línea que separa —por ejemplo— las “acciones coordinadas” de simples campañas políticas es muy difusa.⁸⁷ En este sentido, no es obvio que esta aproximación sea fácil de administrar por parte de las plataformas. Por otro lado, conlleva un costo. Piénsese, por ejemplo, en las restricciones que impuso Whatsapp respecto de la cantidad de veces que determinados mensajes pueden enviarse. La restricción asume que el contenido que se reenvía es problemático, y—por definición—restringe el alcance del mensaje. Ello, en contextos de países autoritarios donde la tecnología de mensajería privada encriptada podría tener algún tipo de rol en el libre flujo de información, es un problema más que una solución.

3. "Policy" y moderación

Las acciones y decisiones de las plataformas tienen que ver con su posicionamiento y sus principios respecto de esta temática, o “policy choices.” Estas elecciones buscan sentar posición sobre el rol que ellas imaginan tener sobre el flujo de información en línea e impactan tanto su ordenamiento interno como sus posicionamientos públicos. Como veremos a continuación, esas posiciones fueron cambiando con el paso del tiempo y tuvieron impactos transversales, sobre todo en la autorregulación y las reglas de moderación. Quizás el cambio mas notorio y memorable para ejemplificar este punto haya sido el cambio de lema de Twitter a partir de 2019: del postulado principio orientador de *speaking truth to power* la empresa paso a “promover una conversación publica sana.”⁸⁸

Hasta 2019 por lo menos prevaleció entre las empresas analizadas una postura de rechazo a adoptar un rol de control sobre la veracidad o falsedad de la información que se tradujo en manifestaciones y posicionamientos públicos y en la adopción de reglas concretas sobre el contenido permisible en sus plataformas. Así, por ejemplo, en Facebook su CEO afirmó que “no chequeamos lo que las

⁸⁷ S. McGregor, «What Even Is “Coordinated Inauthentic Behavior” On Platforms?», WIRED, 17/9/2020, disponible en <https://www.wired.com/story/what-even-is-coordinated-inauthentic-behavior-on-platforms/> Fecha de consulta: 22/septiembre/2020 («Esta ambigüedad y el *enforcement* inconsistente, así como la forma no principista en la que el contenido político es moderado, exacerba las amenazas al proceso electoral, así como la propia capacidad de las plataformas de defenderse de críticos de ambos lados del mostrador»).

⁸⁸ CELE, Comentario en respuesta a los cambios propuestos por Twitter a las reglas sobre contenido deshumanizante, https://www.palermo.edu/Archivos_content/2020/cele/marzo/Internet_y_Derechos_HumanosIII.pdf

personas dicen antes de que lo digan, y francamente no creo que la sociedad quiera que lo hagamos”.⁸⁹ Poco después de la elección presidencial de 2016, su CEO anunció criterios generales respecto de la desinformación.⁹⁰

“Las dos preocupaciones más discutidas el año pasado fueron la diversidad de los puntos de vista a los que accedemos (*filter bubbles*) y la precisión de la información (*fake news*). Me preocupan ambos fenómenos y los estudiamos en profundidad, pero también me preocupa que hay efectos aún más poderosos relacionados con el sensacionalismo y la polarización que llevan a perder un entendimiento común ... Las redes sociales ya proveen puntos de vista más diversos que los medios tradicionales ... Pero nuestro objetivo debe ser que las personas accederán a una foto más completa, no solamente a alternativas diferentes. Debemos ser cuidadosos en cómo hacemos esto. Las investigaciones sugieren que algunas de las ideas más obvias, como mostrar artículos de perspectivas opuestas, de hecho profundizan la polarización porque enmarcan la otra mirada como extranjeras. Una aproximación más efectiva es mostrar un rango de perspectivas, dejar que las personas vean dónde se ubican sus puntos de vista en un espectro y llegar a la conclusión de que lo que piensan es correcto. Con el paso del tiempo, nuestra comunidad va a identificar qué fuentes ofrecen un rango de perspectivas más grande y ese contenido va a aparecer naturalmente más...”

Específicamente sobre la desinformación, Zuckerberg argumentó que es necesario “proceder con cuidado, porque no siempre hay una línea clara entre engaños, sátira y opinión. En una sociedad libre, es importante que las personas tengan el poder de compartir sus opiniones, incluso si otros creen que se equivocan. Nuestra aproximación tendrá que ver menos con prohibir la desinformación, y más con hacer aparecer perspectivas e informaciones adicionales, incluyendo verificadores que disputan la veracidad de determinado ítem”.⁹¹

⁸⁹ “...Zuckerberg además dijo que iba a continuar trabajando con el gobierno de Estados Unidos, iba a profundizar las investigaciones internas, iba a fortalecer el proceso interno de revisión de avisos, duplicar el equipo que trabaja en temas de “integridad de elecciones”, expandir alianzas con autoridades electorales y aumentar la cantidad de información sobre amenazas que comparten con otras empresas tecnológicas. Ver P. Kafka (*Mark Zuckerberg: «freedom Means You Don't Have To Ask For Permission First»*, Vox, 21/09/2017, disponible en <https://www.vox.com/2017/9/21/16346858/mark-zuckberg-facebook-russia-freedom-permission> Fecha de consulta: 4/marzo/2020.)

⁹⁰ M. Zuckerberg, “Building Global Community”, cit.

⁹¹ *Ibid.* En esa línea, por ejemplo, Facebook anunció en abril de 2017 que iba a adoptar una posición más robusta respecto

También en 2017 Facebook lanzó la *Hard Questions Initiative*, que buscó plantear (y responder) preguntas difíciles sobre el rol de las redes sociales en las comunidades políticas democráticas.⁹² En octubre, por ejemplo, reveló información sobre avisos vinculados a Rusia durante la campaña electoral de 2016⁹³ y en noviembre permitió a los usuarios conocer si habían dado “me gusta” a alguna página creada por la *Internet Research Agency*, vinculada a Rusia.⁹⁴ En su documento de *policy*, Facebook define a las “operaciones de información” como las “acciones adoptadas por actores organizados (gubernamentales o no gubernamentales) para distorsionar sentimientos políticos domésticos o foráneos, con frecuencia para lograr un resultado estratégico y/o geopolítico. Estas operaciones pueden usar una combinación de métodos, como noticias falsas, desinformación, o redes de cuentas falsas dirigidas a manipular a la opinión pública (nos referimos a éstas como ‘falsos amplificadores’”.⁹⁵ Google también produjo un breve informe sobre “lo que encontró” en sus plataformas respecto del mismo período⁹⁶ y Twitter anunció que seguiría esos pasos.⁹⁷

Estas posturas pueden haber estado basadas en el paradigma *optimista* respecto de Internet que derivó en la no responsabilidad de los intermediarios por contenidos de terceros. Pero la presión sobre las compañías durante los últimos tres años las llevó a asumir en mayor o menor medida progresiva y reticentemente un rol de “control” en materia de desinformación, sobre todo en contextos electorales. Los cambios en ese sentido son bastante significativos, y muestran a empresas flexibles, con capacidad de adaptar sus políticas a las crecientes presiones

de sus acciones de moderación, para incluir “formas más sutiles de mal uso, incluyendo intentos por manipular el discurso cívico y engañar a la gente”. J. Weedon; W. Nuland; A. Stamos, “Information Operations And Facebook”, cit., pág. 3.

⁹² E. Schrage, *Introducing Hard Questions*, ABOUT FACEBOOK, 15/06/2017, disponible en <https://about.fb.com/news/2017/06/hard-questions/> Fecha de consulta: 4/marzo/2020.

⁹³ E. Schrage, “Hard Questions”, cit.

⁹⁴ Facebook, *Continuing Transparency On Russian Activity*, ABOUT FACEBOOK, 22/11/2017, disponible en <https://about.fb.com/news/2017/11/continuing-transparency-on-russian-activity/> Fecha de consulta: 6/marzo/2020.

⁹⁵ *Ibid.*, pág. 4. Resulta interesante llamar la atención sobre el esfuerzo de Facebook por definir con precisión el fenómeno, que p.ej. distingue a las «operaciones de información» de las «noticias falsas» (“News articles that purport to be factual, but which contain intentional misstatements of fact with the intention to arouse passions, attract viewership, or deceive”) o la «desinformación» (“Inaccurate or manipulated information/content that is spread intentionally. This can include false news, or it can involve more subtle methods, such as false flag operations, feeding inaccurate quotes or stories to innocent intermediaries, or knowingly amplifying biased or misleading information. Disinformation is distinct from misinformation, which is the inadvertent or unintentional spread of inaccurate information without malicious intent”).

⁹⁶ Google, «What We Found». Google, Palo Alto, California. 30 de octubre de 2017.

⁹⁷ T. Romm, *Twitter Is Exploring How To Notify Its Users That They Saw Russian Propaganda During The 2016 Election*, Vox, 17/01/2018, disponible en <https://www.vox.com/2018/1/17/16901428/twitter-notify-users-russia-trolls-2016-election> Fecha de consulta: 6/marzo/2020; Twitter public policy, *Update On Twitter's Review Of The 2016 Us Election*, 19/01/2018, disponible en https://blog.twitter.com/en_us/topics/company/2018/2016-election-update.html Fecha de consulta: 9/marzo/2020.

y jugar un rol más activo y transparente en el control de la información que fluye por sus plataformas.

Por ejemplo, Facebook parece haberse esforzado por defender cierta visión sobre la libertad de expresión que por lo menos en teoría, seguiría justificando no tener que ejercer este rol.⁹⁸ Sin embargo progresivamente—y creemos que como consecuencia de la creciente presión sobre la empresa—también fue adoptando cambios en su postura. Por ejemplo, en marzo de 2018 Facebook expandió sus políticas de verificación de contenidos a las fotos y los vídeos.⁹⁹ Si hasta ese momento verificaba “noticias”, la decisión de expandir sus esfuerzos de verificación en colaboración con terceros supuestamente calificados para ello es una decisión en tensión con el posicionamiento más general de la compañía. En octubre de 2018 también anunció esfuerzos específicos para luchar contra la desinformación que buscaba “suprimir” el voto,¹⁰⁰ algo que ratificó en septiembre de 2020 de cara a las elecciones de ese año.¹⁰¹

Durante 2019, estos cambios de *policy* respecto de los anuncios durante campañas políticas parecieron hacerse más estrictos. En enero, Facebook anunció reglas más duras para lidiar con avisos publicitarios.¹⁰² Google resolvió dejar de emitir publicidad política en Canadá por reglas de transparencia que serían “muy difíciles de cumplir”;¹⁰³ en marzo de 2019 decidió relanzar su *Ad Library*¹⁰⁴; y en abril de 2020 anunció un sistema de verificación de identidad para algunos de sus avisos.¹⁰⁵

⁹⁸ E. Schrage, *Hard Questions: Russian Ads Delivered To Congress*, ABOUT FACEBOOK, 02/10/2017, disponible en <https://about.fb.com/news/2017/10/hard-questions-russian-ads-delivered-to-congress/> Fecha de consulta: 6/marzo/2020 («...must also take seriously the crucial place that free political speech occupies around the world in protecting democracy and the rights of those who are in the minority, who are oppressed or who have views that are not held by the majority or those in power»); K. Swisher, *Full Transcript: Facebook CEO Mark Zuckerberg On Recode Decode*, Vox, 18/07/2018, disponible en <https://www.vox.com/2018/7/18/17575158/mark-zuckerberg-facebook-interview-full-transcript-kara-swisher> Fecha de consulta: 11/marzo/2020.

⁹⁹ G. Rosen, *Hard Questions: What Is Facebook Doing To Protect Election Security?*, ABOUT FACEBOOK, 29/03/2018, disponible en <https://about.fb.com/news/2018/03/hard-questions-election-security/> Fecha de consulta: 9/marzo/2020.

¹⁰⁰ J. Menn, «Exclusive: Facebook To Ban Misinformation On Voting In Upcoming U.s. Elections», REUTERS, 16/10/2018, disponible en <https://www.reuters.com/article/us-facebook-election-exclusive-idUSKCN1MP2G9> Fecha de consulta: 11/marzo/2020.

¹⁰¹ About Facebook, *New Steps To Protect The Us Elections*, ABOUT FACEBOOK, 03/09/2020, disponible en <https://about.fb.com/news/2020/09/additional-steps-to-protect-the-us-elections/> Fecha de consulta: 3/septiembre/2020.

¹⁰² *Youtube Just Demonetized Anti-vax Channels*, BUZZFEED NEWS, 22/02/2019, disponible en <https://www.buzzfeednews.com/article/carolineodonovan/youtube-just-demonetized-anti-vax-channels> Fecha de consulta: 11/marzo/2020.

¹⁰³ T. Cardoso, «Google To Ban Political Ads Ahead Of Federal Election, Citing New Transparency Rules», 4/3/2019, disponible en <https://www.theglobeandmail.com/politics/article-google-to-ban-political-ads-ahead-of-federal-election-citing-new/> Fecha de consulta: 11/marzo/2020.

¹⁰⁴ S. Shukla, *A Better Way To Learn About Ads On Facebook*, ABOUT FACEBOOK, 28/03/2019, disponible en <https://about.fb.com/news/2019/03/a-better-way-to-learn-about-ads/> Fecha de consulta: 11/marzo/2020.

¹⁰⁵ J. Canfield, *Increasing Transparency Through Advertiser Identity Verification*, GOOGLE, 23/04/2020, disponible en <https://>

La presión sobre las empresas también derivó en acciones de revisión de actividades sospechosas. En septiembre de 2017 Facebook anunció que había identificado gastos por más de 100,000 dólares de cuentas falsas provenientes de Rusia (que canceló, siguiendo su política de no permitir cuentas “no auténticas”).¹⁰⁶ En octubre de 2018, Twitter publicó información sobre “potenciales operaciones de información”.¹⁰⁷ Y Facebook dio información sobre algunos casos de noticias falsas, en las que reportó qué detectaron, cómo verificaron la información, y—quizás de manera más relevante—cómo lo detectaron.¹⁰⁸ Dicha revelación sugiere que las herramientas preferidas de Facebook para identificar información falsa son *machine learning* y reportes de verificadores externos.¹⁰⁹

Todas estas acciones dan cuenta de un cambio en la forma en que las empresas se ven a sí mismas y se muestran a sus diferentes audiencias.

Los cambios de policy impactan directamente en la definición e implementación de las políticas de contenidos, las *community guidelines* o reglas internas mediante las cuales las compañías toman decisiones de moderación.¹¹⁰ En este caso, se trata de cambios a los contenidos que—para las propias plataformas—son aceptables o no.

En abril de 2017, por ejemplo, Facebook emitió un documento en el que definió algunos conceptos clave vinculados a las “operaciones de información” que impactaron en sus políticas internas.¹¹¹ En junio de 2019, Twitter actualizó sus reglas y las simplificó para que sean más fáciles de entender.¹¹² Hacia septiembre de 2019, sin embargo, Twitter y Facebook anunciaron cambios más radicales que revelan su

blog.google/products/ads/advertiser-identity-verification-for-transparency/ Fecha de consulta: 27/agosto/2020.

¹⁰⁶ A. Stamos, *An Update On Information Operations On Facebook*, ABOUT FACEBOOK, 06/09/2017, disponible en <https://about.fb.com/news/2017/09/information-operations-update/> Fecha de consulta: 4/marzo/2020.

¹⁰⁷ V. Gadde; Y. Roth, *Enabling Further Research Of Information Operations On Twitter*, TWITTER BLOG, 17/10/2018, disponible en https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html Fecha de consulta: 11/marzo/2020.

¹⁰⁸ A. Woodford, *The Hunt For False News*, ABOUT FACEBOOK, 19/10/2018, disponible en <https://about.fb.com/news/2018/10/inside-feed-hunt-false-news-october-2018/> Fecha de consulta: 11/marzo/2020.

¹⁰⁹ *Ibid.*

¹¹⁰ Pero estos cambios no necesariamente se expresan, automáticamente, en cambios formales a sus reglas. Ver sobre este punto, T. Gillespie (*Custodians of the Internet*, cit.), 66 (“Though Facebook’s policy was instituted in December 2016, as of October 2017, no new wording has been added to the community standards document that could reasonably refer to this change. One could argue that Facebook does not prohibit fake news, it only labels it; another possibility is that in some cases policies are instituted, but their addition to community guidelines documents lags behind”). Las reglas al momento de escribir este reporte (julio de 2020) incluyen a las “noticias falsas” como contenido prohibido. Ver https://www.facebook.com/communitystandards/false_news. Sobre esta cuestión, ver el proyecto *Letra Chica* del CELE y Linterna Verde, en <https://letrachica.digital/>

¹¹¹ J. Weedon; W. Nuland; A. Stamos, “Information Operations And Facebook”, cit.

¹¹² D. Harvey, *Making Our Rules Easier To Understand*, TWITTER BLOG, 07/06/2019, disponible en https://blog.twitter.com/en_us/topics/company/2019/rules-refresh.html Fecha de consulta: 11/marzo/2020.

propia aproximación a la cuestión y—en menor medida—la magnitud del desafío. Así, Twitter anunció la suspensión total de la publicidad política, bajo la premisa de que el “alcance” de los mensajes políticos se tiene que “ganar, y no comprar”.¹¹³

“A finales del año pasado, Twitter tomó la decisión de prohibir en todo el mundo la promoción de contenido de carácter político. Tomamos esta decisión con base en nuestra creencia de que el alcance de los mensajes políticos se debe ganar, no comprar. Definimos el contenido de carácter político como aquel que hace referencia a un candidato, partido político, funcionario gubernamental electo o designado, elección, referéndum, medida sometida a votación, ley, normativa, directiva o fallo judicial. Los anuncios que contengan referencias a contenido de carácter político, incluidas peticiones de votos, solicitudes de apoyo financiero y promoción a favor o en contra de los tipos de contenido de carácter político mencionados anteriormente, quedan prohibidos en virtud de esta política. Tampoco permitiremos anuncios de ningún tipo de parte de candidatos, partidos políticos o funcionarios del gobierno elegidos o designados. Cualquier persona que use la plataforma puede reportar anuncios que vayan en contra de esta política. El mecanismo de reporte a través de la propia aplicación está disponible globalmente y nuestros usuarios y usuarias pueden reportar a través de este canal. Nosotros estamos trabajando junto a las autoridades electorales alrededor del mundo para asegurar un diálogo abierto que abone a la identificación de anuncios violatorios de esta política”.¹¹⁴

Por su parte, Facebook anunció que ya no verificaría más información política, una acción sorpresiva, que implicó un cambio de dirección y que la propia compañía explicó en términos de libertad de expresión.¹¹⁵ En septiembre de 2020

¹¹³ (5) Jack On Twitter: «we've Made The Decision To Stop All Political Advertising On Twitter Globally. We Believe Political Message Reach Should Be Earned, Not Bought. Why? A Few Reasons... » / Twitter, TWITTER, 30/10/2019, disponible en <https://twitter.com/jack/status/1189634360472829952> Fecha de consulta: 11/marzo/2020.

¹¹⁴ Hugo Rodríguez Nicolat, ‘Comentarios de Twitter a La Encuesta de Investigación: “Desinformación En Internet: Acciones y Reacciones” (2020, en archivo con los autores)., cit., pág. 4.

¹¹⁵ N. Clegg, Facebook, *Elections And Political Speech*, ABOUT FACEBOOK, 24/09/2019, disponible en <https://about.fb.com/news/2019/09/elections-and-political-speech/> Fecha de consulta: 11/marzo/2020 («We don't believe, however, that it's an appropriate role for us to referee political debates and prevent a politician's speech from reaching its audience and being subject to public debate and scrutiny. That's why Facebook exempts politicians from our third-party fact-checking program. We have had this policy on the books for over a year now, posted publicly on our site under our eligibility guidelines»).

introdujo otro cambio significativo: siguiendo a Twitter, Facebook anunció que no aceptarían avisos políticos en la semana previa a la elección.¹¹⁶

Tres sucesos recientes han impulsado a las plataformas a volver a modificar sus aproximaciones al fenómeno de la desinformación. Se trata de la pandemia global causada por el SARS-CoV-2 y la enfermedad del COVID-19, las protestas del movimiento de *Black Lives Matter* de 2020 y las elecciones presidenciales en los Estados Unidos. Todos estos eventos mostraron la dificultad de moderar contenidos de una manera satisfactoria para todos los actores que demandan mejoras en ese sentido.¹¹⁷

La pandemia causada por el SARS-CoV-2 disparó otra fuente de presión, en este caso motorizada por las autoridades sanitarias globales, regionales y nacionales. Y las plataformas parecen haber reaccionado al nuevo escenario publicando información confiable sobre la pandemia, de manera proactiva.¹¹⁸

“Por ejemplo, Facebook estableció un ‘Centro de Información sobre el COVID-19’, compartiendo información oficial y otros medios creíbles. Este centro promovió diversos contenidos sobre la pandemia, incluyendo algunos que eran confusos. YouTube empezó a dirigir a las personas que veían videos relacionados con el virus a ‘conseguir los últimos datos del coronavirus’ de fuentes oficiales como la Organización Mundial de la Salud y gobiernos nacionales. Twitter, Instagram y TikTok hicieron cosas similares”.¹¹⁹

¹¹⁶ About Facebook, “New Steps To Protect The Us Elections”, cit.

¹¹⁷ Sobre este punto, ver A. Wiener (*Trump, Twitter, Facebook, And The Future Of Online Speech*, cit.); D. Alba; K. Conger; R. Zhong (“Twitter Adds Warnings To Trump And White House Tweets, Fueling Tensions”, cit.); M. Murphy («Facebook Shouldn’t Be The Arbiter Of Truth,» Zuckerberg Tells Fox News, MARKETWATCH, 28/05/2020, disponible en <https://www.marketwatch.com/story/facebook-shouldnt-be-the-arbiter-of-truth-zuckerberg-tells-fox-news-2020-05-27> Fecha de consulta: 15/julio/2020). En junio de 2020 Facebook también anunció que empezaría a “etiquetar” contenidos de políticos, en la misma línea que Twitter. Cf. M. Zuckerberg (*Mark Zuckerberg’s Announcement*, ZUCKERBERG’S FACEBOOK WALL, 26/06/2020, disponible en <https://www.facebook.com/zuck/posts/10112048980882521> Fecha de consulta: 15/julio/2020.) (“A handful of times a year, we leave up content that would otherwise violate our policies if the public interest value outweighs the risk of harm. Often, seeing speech from politicians is in the public interest, and in the same way that news outlets will report what a politician says, we think people should generally be able to see it for themselves on our platforms. We will soon start labeling some of the content we leave up because it is deemed newsworthy, so people can know when this is the case. We’ll allow people to share this content to condemn it, just like we do with other problematic content, because this is an important part of how we discuss what’s acceptable in our society – but we’ll add a prompt to tell people that the content they’re sharing may violate our policies. To clarify one point: there is no newsworthiness exemption to content that incites violence or suppresses voting. Even if a politician or government official says it, if we determine that content may lead to violence or deprive people of their right to vote, we will take that content down. Similarly, there are no exceptions for politicians in any of the policies I’m announcing here today”)

¹¹⁸ J. D’Urso, *How The Coronavirus Pandemic Is Changing Social Media*, REUTERS INSTITUTE FOR THE STUDY OF JOURNALISM, 06/07/2020, disponible en <https://reutersinstitute.politics.ox.ac.uk/risj-review/how-coronavirus-pandemic-changing-social-media> Fecha de consulta: 10/julio/2020.

¹¹⁹ *Ibid.*

Este tipo de prácticas no ha estado exenta de importantes desafíos. Por un lado, no todos los gobiernos son fuente confiable de noticias o información, incluso en contextos como el de la pandemia. Por otro lado, como señala D’Urso, las afirmaciones gubernamentales no necesariamente reciben los “clicks” necesarios para llegar a la población, especialmente cuando compiten con contenidos que—como los que usualmente conforman las campañas de desinformación—apelan a reacciones emocionales de los lectores.¹²⁰ Como hemos visto, recomendaciones sobre uso de medicamentos no comprobados como efectivos, remedios caseros o—más sencillamente—la subestimación de los riesgos de contagio fueron usuales en las Américas en los últimos meses. La desinformación sobre la pandemia y las características especiales que ella supuestamente tiene (riesgos más claros, supuesta facilidad para demostrar la falsedad de las afirmaciones) habrían llevado a las plataformas a “hacer más” en relación a este tipo de desinformación que con respecto a las cuestiones electorales.¹²¹

La presión para *actuar* sin embargo no se limitó a cuestiones de salud y siguió creciendo a lo largo del 2020. En Julio de 2020, por ejemplo, poderosos anunciantes anunciaron un boicot a Facebook para que tome acciones más decisivas respecto de discursos problemáticos, en especial el “discurso de odio”.¹²² La suspensión—por ahora, *sine die*—de las cuentas de Donald Trump de las tres plataformas marca el cambio más profundo y radical hasta el momento.

Indudablemente uno de los principales problemas de la moderación de contenidos y la autorregulación tiene que ver con la maleabilidad de los criterios y el hecho, significativo, de que muchas veces esos criterios no se expresan de manera precisa en los términos y condiciones o las *community guidelines*. Para abordar esta dificultad, distintas iniciativas han procurado darle seguimiento a estos complejos documentos legales. Por ejemplo, el CELE y Linterna Verde han lanzado recientemente *LetrChica*, la primera iniciativa que sigue esta cuestión en

¹²⁰ Cf. V. Bakir; A. McStay, «Fake News And The Economy Of Emotions: Problems, Causes, Solutions», *DIGITAL JOURNALISM*, vol. 6, 2, 2018, Disponible en <https://www.tandfonline.com/doi/full/10.1080/21670811.2017.1345645> («...the fakes news problem concerns the *economics of emotion* ... emotions are leveraged to generate attention and viewing time, which converts to advertising revenue...»).

¹²¹ J. D’Urso, *How The Coronavirus Pandemic Is Changing Social Media*, cit. («During the pandemic, social media companies have shown some signs of going further than before when it comes to removing content...»).

¹²² «Mark Zuckerberg: Advertisers’ Boycott Of Facebook Will End “soon Enough”», *THE GUARDIAN*, 2/7/2020, disponible en <https://www.theguardian.com/technology/2020/jul/02/mark-zuckerberg-advertisers-boycott-facebook-back-soon-enough>
Fecha de consulta: 15/julio/2020.

4. Acciones de transparencia y relaciones públicas

La evolución de las políticas, las normas de autorregulación y la moderación de contenidos vino acompañada de grandes presiones en materia de transparencia. Las tres compañías bajo análisis publican “reportes de transparencia” que dan cuenta de una intensa actividad de moderación, aunque el nivel de detalle que ofrecen y los formatos en los que la información se brinda dificultan el análisis de la información.¹²⁴

En materia electoral, y como medida alternativa a la moderación, las empresas adoptaron progresivamente medidas para incrementar los niveles de transparencia sobre los avisos publicitarios:¹²⁵ señalar el origen del aviso, pero ofrecer—además—información de contexto, alcance general de la campaña, montos invertidos, etcétera.¹²⁶ Facebook anunció cambios y expansiones a esta política en diversas oportunidades.¹²⁷ Twitter también anunció una política similar casi simultáneamente, a través de la creación del *Advertising Transparency Center*.¹²⁸ Google, por su parte, adoptó una política de transparencia más estricta de cara a las elecciones de medio término de 2018¹²⁹ y lanzó un sitio web que contiene información agregada sobre avisos políticos, restringido a Estados Unidos, la

¹²³ Disponible en <https://letrachica.digital/>

¹²⁴ Ver <https://transparency.facebook.com/>, <https://transparency.twitter.com/>, <https://transparencyreport.google.com/?hl=es>

¹²⁵ P. Kafka, *Mark Zuckerberg*, cit.

¹²⁶ Esta acción tuvo diversas consecuencias, ya que Facebook extendió su política a todos los avisos, y no sólo a los avisos de naturaleza política en octubre de 2017. Cf. R. Goldman (*Update On Our Advertising Transparency And Authenticity Efforts*, ABOUT FACEBOOK, 27/10/2017, disponible en <https://about.fb.com/news/2017/10/update-on-our-advertising-transparency-and-authenticity-efforts/> Fecha de consulta: 6/marzo/2020.)

¹²⁷ R. Goldman, *Making Ads And Pages More Transparent*, About Facebook, 06/04/2018, disponible en <https://about.fb.com/news/2018/04/transparent-ads-and-pages/> Fecha de consulta: 9/marzo/2020 (en abril de 2018); K. Harbath, *Updates To Ads About Social Issues, Elections Or Politics In The Us*, About Facebook, 28/08/2019, disponible en <https://about.fb.com/news/2019/08/updates-to-ads-about-social-issues-elections-or-politics-in-the-us/> Fecha de consulta: 11/marzo/2020 (en mayo, referida a elecciones); R. Leathern, *Shining A Light On Ads With Political Content*, About Facebook, 24/05/2018, disponible en <https://about.fb.com/news/2018/05/ads-with-political-content/> Fecha de consulta: 11/marzo/2020; K. Walker, *Supporting Election Integrity Through Greater Advertising Transparency*, Google, 04/05/2018, disponible en <https://blog.google/outreach-initiatives/public-policy/supporting-election-integrity-through-greater-advertising-transparency/> Fecha de consulta: 9/marzo/2020.

¹²⁸ B. Falck, *New Transparency For Ads On Twitter*, TWITTER, 27/10/2017, disponible en https://blog.twitter.com/en_us/topics/product/2017/New-Transparency-For-Ads-on-Twitter.html Fecha de consulta: 6/marzo/2020. El centro se encuentra disponible en <https://ads.twitter.com/transparency>.

¹²⁹ K. Walker, *Supporting Election Integrity Through Greater Advertising Transparency*, cit.

India, y el Reino Unido y la Unión Europea.¹³⁰

En la misma línea, establecieron lazos de colaboración con autoridades electorales, ofreciendo diálogo en tiempo real y cierta apertura sobre lo que estaba ocurriendo en sus plataformas. Facebook anunció la expansión de esta política en septiembre de 2017¹³¹ y mostró resultados de esa colaboración en relación a las elecciones alemanas.¹³² De manera similar, Twitter destacó que las alianzas con las autoridades electorales son una constante de la compañía para reaccionar ante escenarios electorales específicos.¹³³ Facebook anunció “medidas especiales” de cara a las elecciones de 2020.¹³⁴

Durante este periodo, además, las empresas han sido llamadas a explicar públicamente sus negocios, modelos, políticas y normas en distintos foros y han sido convocadas explícitamente a encuentros formales con funcionarios públicos (p. Ej. Estados Unidos, Reino Unido). Las citaciones del Congreso de los Estados parecen ser especialmente relevantes. En septiembre de 2017, *The Hill* reportó que Facebook, Google y Twitter serían llamados a testificar ante el comité de Inteligencia del Senado por interferencia rusa en las elecciones de 2016.¹³⁵ En ese marco, Twitter anunció

¹³⁰ Google, *Publicidad Política En Google – Informe De Transparencia De Google*, GOOGLE INFORME DE TRANSPARENCIA, 15/08/2018, disponible en <https://transparencyreport.google.com/political-ads/region/US> Fecha de consulta: 11/marzo/2020.

¹³¹ P. Kafka, *Mark Zuckerberg*, cit.

¹³² R. Allan, *Update On German Elections*, ÜBER FACEBOOK, 27/09/2017, disponible en <https://about.fb.com/de/news/2017/09/update-zu-den-wahlen/> Fecha de consulta: 4/marzo/2020 («Facebook co-operated closely with German authorities, such as the Federal Office for Information Security (BSI). We provided training for members of Parliament and candidates on online security issues. We also set up a dedicated reporting channel for the BSI for issues related to security and authenticity in the context of the federal elections»).

¹³³ Twitter public policy, *Update*, cit. («We engage with national election commissions regularly and consistently bolster our security and agent review coverage during key moments of election cycles around the world. We will continue to do this, and expand our outreach so that we've got strong, clear escalation processes in place for all potentialities during major elections and global political events»); Hugo Rodríguez Nicolat, «Comentarios de Twitter a La Encuesta de Investigación: “Desinformación En Internet: Acciones y Reacciones”» (2020, en archivo con los autores), cit., pág. 3 («En este sentido, de manera similar a los modelos que hemos puesto en práctica para elecciones recientes alrededor del mundo, como es el caso de Estados Unidos, Brasil y México, hemos implementado un equipo interno e interdisciplinario para liderar nuestros esfuerzos en materia de integridad electoral. A través de nuestras propias herramientas internas, este equipo trabaja proactivamente para evaluar tendencias, atender casos que son escalados por nuestros aliados en la materia e identificar posibles amenazas provenientes de actores maliciosos. Respecto de colaboraciones específicas con autoridades electorales y actores similares en Latinoamérica, les invitamos a leer nuestro blog referente a las Elecciones en México a manera de ejemplo. Esfuerzos similares fueron de igual forma implementados en Colombia y Argentina. Asimismo, dada la similitud entre esta pregunta y la pregunta número seis, en la respuesta a esta última ahondaremos más al respecto»).

¹³⁴ About Facebook, «New Steps To Protect The Us Elections», cit.; G. Rosen, *Helping To Protect The 2020 Us Elections*, ABOUT FACEBOOK, 21/10/2019, disponible en <https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/> Fecha de consulta: 11/marzo/2020.

¹³⁵ J. Byrnes (*Senate Panel Invites Facebook, Google To Testify In Russia Probe*, THE HILL, 27/09/2017, disponible en <https://thehill.com/homenews/senate/352743-senate-panel-invites-facebook-to-testify> Fecha de consulta: 7/mayo/2020.). En ese momento Facebook anunció que cooperaría con las autoridades. Ver C. Stretch (*Facebook To Provide Congress With Ads Linked To Internet Research Agency*, ABOUT FACEBOOK, 21/09/2017, disponible en <https://about.fb.com/news/2017/09/provid->

encuentros y señaló obstáculos para lidiar con cuentas falsas y *bots*.¹³⁶ En noviembre de 2017 las empresas anunciaron que colaborarían con el Reino Unido para investigar posibles interferencias rusas en el referéndum por la permanencia de ese país en la Unión Europea de 2016.¹³⁷ En septiembre de 2018, Sheryl Sandberg de Facebook y Jack Dorsey de Twitter testificaron ante el Comité de Inteligencia del Senado de EEUU nuevamente.¹³⁸ Resulta interesante señalar que hay un subtexto de cuestionamiento a los avisos direccionados que empuja a las compañías a su defensa,¹³⁹ posicionamientos que se ven—por ejemplo—en los “principios” que las plataformas sostienen respecto de sus prácticas publicitarias.¹⁴⁰

Finalmente y de forma paralela al desarrollo de instancias de seguimiento estatal, las propias empresas han desarrollado acciones y creado procedimientos y mecanismos con distintos niveles de apertura a terceros (especialistas, ONGs, académicos, etc) para generar políticas y criterios de implementación que satisfagan a usuarios y reguladores. En abril de 2018, Facebook publicó sus criterios internos de moderación y empezó a celebrar reuniones con partes interesadas sobre las mismas; y expandió su procedimiento interno de apelación sobre sus decisiones.¹⁴¹ Para entonces, Twitter ya había lanzado su *Trust and Safety Council* y recientemente anunció una expansión de sus facultades en línea con el cuerpo anunciado por Facebook.¹⁴² A fines de 2018 también Facebook anunció la crea-

ing-congress-with-ads-linked-to-internet-research-agency/ Fecha de consulta: 6/marzo/2020.)

¹³⁶ Wired, «Twitter Will Meet With Senate Intelligence Committee On Russia», WIRED, 28/9/2017, disponible en <https://www.wired.com/story/twitter-senate-committee-russia-bots/> Fecha de consulta: 7/mayo/2020.

¹³⁷ M. D. Stefano, *Facebook And Twitter Say They'll Cooperate With The Uk Inquiry Into Russian Meddling In Brexit*, BUZZFEED, 29/11/2017, disponible en <https://www.buzzfeed.com/markdistefano/facebook-and-twitter-say-theyre-ready-to-co-operate-with> Fecha de consulta: 6/marzo/2020.

¹³⁸ US Senate Intelligence Committee, *Hearings | Intelligence Committee*, US SENATE, 05/09/2018, disponible en <https://www.intelligence.senate.gov/hearings/open-hearing-foreign-influence-operations%E2%80%99-use-social-media-platforms-company-witnesses> Fecha de consulta: 7/mayo/2020.

¹³⁹ E. Schrage, “Hard Questions”, cit. («These are worthwhile uses of ad targeting because they enable people to connect with the things they care about. But we know ad targeting can be abused, and we aim to prevent abusive ads from running on our platform. To begin, ads containing certain types of targeting will now require additional human review and approval»).

¹⁴⁰ Ver, p.ej., los principios de Facebook en R. Goldman, *Our Advertising Principles*, ABOUT FACEBOOK, 27/11/2017, disponible en <https://about.fb.com/news/2017/11/our-advertising-principles/> Fecha de consulta: 6/marzo/2020.

¹⁴¹ M. Bickert (*Publishing Our Internal Enforcement Guidelines And Expanding Our Appeals Process*, ABOUT FACEBOOK, 24/04/2018, disponible en <https://about.fb.com/news/2018/04/comprehensive-community-standards/> Fecha de consulta: 9/marzo/2020.). Los *community standards* se pueden encontrar, actualizados, en <https://www.facebook.com/community-standards/>. En mayo de 2018, Facebook publicó un reporte sobre casos que buscó mostrar cómo sus criterios operan en la práctica. Ver G. Rosen (*Facebook Publishes Enforcement Numbers For The First Time*, ABOUT FACEBOOK, 15/05/2018, disponible en <https://about.fb.com/news/2018/05/enforcement-numbers/> Fecha de consulta: 9/marzo/2020.).

¹⁴² N. Pickles, *Strengthening Our Trust And Safety Council*, TWITTER BLOG, 13/12/2019, disponible en https://blog.twitter.com/en_us/topics/company/2019/strengthening-our-trust-and-safety-council.html Fecha de consulta: 27/agosto/2020.

ción de un consejo de notables que ayudaría a la compañía a adoptar buenas decisiones en materia de moderación, compatibles con los principios de libertad de expresión que públicamente Zuckerberg subrayó en diversas posibilidades. El *Oversight Board* fue finalmente presentado públicamente en mayo de 2020.¹⁴³

Este tipo de acciones parecen responder a las demandas crecientes de “rendición de cuentas” en materia de moderación de contenidos y es un paso en la dirección de convertir a las grandes plataformas en “foros públicos”.

C. Análisis: buenas prácticas, dilemas y desafíos

El fenómeno de la desinformación emergió como problema después de la elección presidencial de 2016 en los Estados Unidos. Las grandes plataformas de Internet han tenido que lidiar con ello desde entonces. En su informe anterior, el CELE detectó que a diciembre de 2017 las plataformas habían tomado acciones que se encontraban “en fase preliminar”.¹⁴⁴

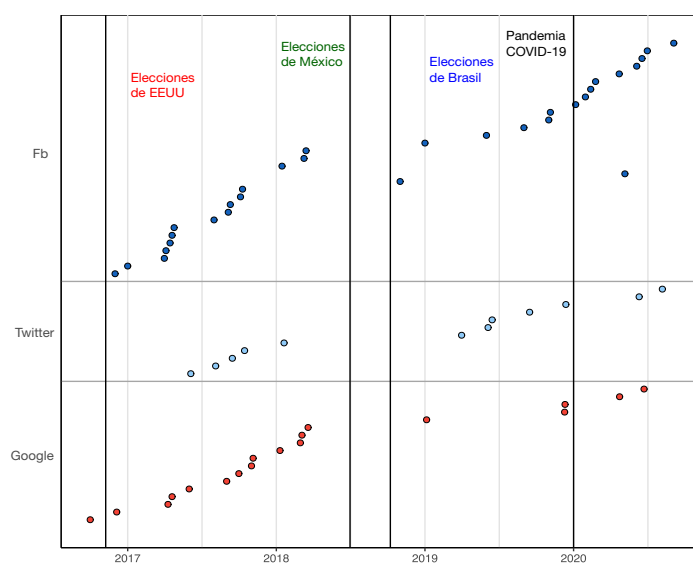


Figura 3 - Timeline de acciones

Vale la pena recordarlas para tener un panorama completo de los hallazgos que este informe busca actualizar.

¹⁴³ (*Establishing Structure And Governance For An Independent Oversight Board*, ABOUT FACEBOOK, 17/09/2019, disponible en <https://about.fb.com/news/2019/09/oversight-board-structure/> Fecha de consulta: 11/marzo/2020.); ver <https://www.oversightboard.com/>

¹⁴⁴ C. Cortés; L. Isaza, “Noticias falsas en Internet: La estrategia para combatir la desinformación”, cit., pág. 26.

- *Facebook*. Actuaba sobre la base de una definición de “operaciones informativas” las acciones desarrolladas por actores organizados para distorsionar el sentimiento político nacional o extranjero”, y se dividen en noticias falsas, desinformación y cuentas falsas. El informe de 2017 identificaba dos tipos de acciones: las positivas, enfocadas a la concientización de usuarios, y las negativas que eran soluciones “técnicas” al desafío de la desinformación. Estas últimas incluían al (a) reporte, verificación de hechos y advertencias; (b) cambios en el newsfeed y medidas para contrarrestar la falsa amplificación.
- *Google*. Además de acciones de concientización, el informe daba cuenta de cambios en los servicios de búsqueda, especialmente vinculados a destacar contenido “de calidad” y servicios de verificación. El informe señalaba, sin embargo, que “Google se limita a resaltar las verificaciones hechas por terceros a partir de ciertos criterios, incluso si éstas han alcanzado conclusiones diferentes”.¹⁴⁵ El punto parece significativo, porque una revisión del sistema de búsqueda actual desde la Argentina no permitió verificar que algún sistema de clasificación que resalte la verificación de contenidos tenga un lugar especial en los resultados de búsqueda usuales. Por otro lado, el informe también identificó la posibilidad de reportar las predicciones, opción que se encuentra disponible en América Latina pero no vinculada a la desinformación.
- *Twitter*. Las acciones identificadas en 2017 incluían la reducción de visibilidad de *tweets* y de posibles cuentas de spam mientras son sometidos/as a investigación, la suspensión de cuentas y medidas para detectar aplicaciones que abusan de su API.

A la vez, el informe reportaba desafíos y posibles soluciones.¹⁴⁶ Nuevamente, cabe consignarlas: señalaba problemas de escala y tiempo y señalaba que la desinformación no podía ser abordada sólo con mecanismos automatizados, por lo que requeriría una cantidad de recursos humanos inviable. Asimismo, el informe destacaba que tomar “buenas” decisiones de moderación en este frente demandaba de la intervención de personas con la capacidad de juicio necesaria para dictaminar, con cierto grado de certidumbre, que efectivamente se está ante una campaña de desinformación problemática. Además, el informe señaló que esas estructuras son difíciles de imaginar en mercados secundarios como los de

¹⁴⁵ *Ibid.*, pág. 17.

¹⁴⁶ *Ibid.*, págs. 21-24.

América Latina. El informe también denunciaba la posibilidad de que algunas de las soluciones propuestas fueran pasibles de las dinámicas de “consecuencias inesperadas”: que—por ejemplo—la verificación de información tuviera el efecto contrario al deseado, como el de reforzar las creencias erróneas o las teorías de la conspiración de las que muchas veces esas creencias se alimentan.

Finalmente, cabe recordar un punto adicional: el informe de 2017 afirmaba que “la elaboración de este documento tuvo una dificultad constante: entender si la gran cantidad de decisiones y acciones cubiertas en medios de comunicación y anunciadas por las empresas eran efectivamente implementadas y en qué medida”.¹⁴⁷ Esta dificultad subsistió durante la elaboración de este documento de seguimiento: de las 61 acciones más relevantes identificadas y reseñadas en este documento, la implementación efectiva de 28 de ellas no pudo ser verificada y la mayor parte de las acciones verificadas pertenecen a las acciones más visibles, como las de apoyo al periodismo, alianzas con verificadores, etcétera (Figura 4).

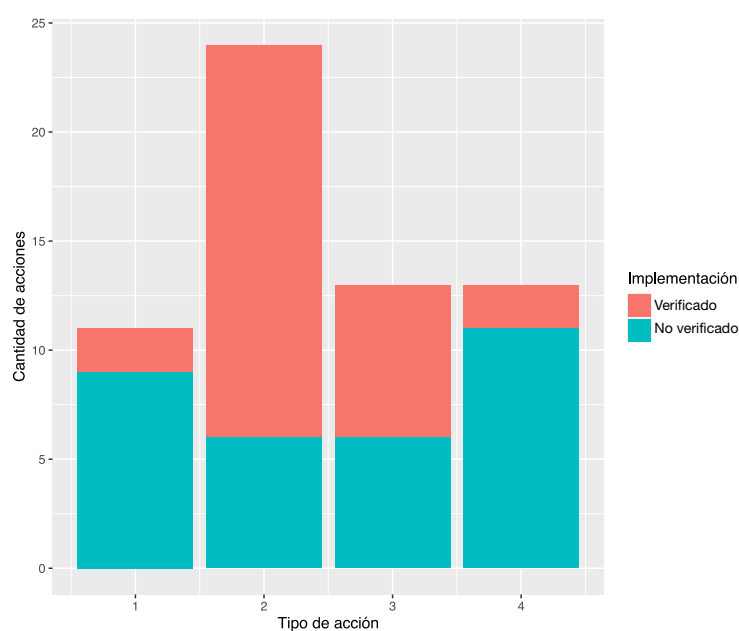


Figura 4 - Tipos de acciones y estado de implementación

El análisis de lo ocurrido en los últimos años sugiere que los patrones identificados en el informe anterior continúan vigentes, pero algunas tendencias novedosas también parecen haberse hecho presentes. Así, por caso, la creación de

¹⁴⁷ *Ibid.*, pág. 23.

organismos internos encargados de incidir en las políticas de moderación aparece como una de las novedades más significativas. El *Oversight Board* de Facebook y—en menor medida—el *Trust and Safety Council* de Twitter son intentos novedosos de parte de las plataformas de aumentar los niveles de rendición de cuentas respecto de sus acciones y políticas de moderación de contenidos, aunque ambas acciones parecen seguir la lógica del “foro público” que lentamente parece imponerse respecto de este tipo de empresas.

Por otro lado, en los últimos tiempos parecen haberse profundizado ciertas tendencias: además de las acciones *seguras* de concientización y promoción de más información, las plataformas aquí reveladas han apostado a sus prácticas de “contextualización” de la información que presentan a los usuarios, cada vez más gracias a algoritmos de recomendaciones de contenidos. La práctica del etiquetado parece haberse expandido: además del etiquetado vinculado a la verificación de contenidos, tanto Twitter como Facebook han etiquetado publicaciones de importantes actores públicos por contener información falsa, medios gubernamentales de países extranjeros, etcétera. Las empresas parecen seguir apostando a las soluciones de inteligencia artificial para sus políticas de moderación: un algoritmo lo suficientemente bien entrenado podría identificar información falsa y actuar de alguna manera, ya sea etiquetándola como tal, estableciendo algún tipo de advertencia previa, disminuyendo su capacidad de circulación en la red, etcétera. Esta confianza en los algoritmos sin embargo podría ser excesiva. Como lo reconocen las propias compañías, la línea que separa a la desinformación de la sátira o la opinión protegida por la libertad de expresión es muy delgada y—por el momento—no parece posible que una solución puramente técnica pueda ser efectiva por sí sola, a pesar de que como consecuencia de la pandemia las plataformas han acelerado la adopción de este tipo de herramientas.¹⁴⁸

Finalmente, algunas tendencias generales parecen estar consolidándose. Las compañías analizadas en el presente estudio—especialmente Facebook y Twitter—parecen haber abrazado un ejercicio robusto de sus facultades de moderación. Las últimas acciones identificadas parecen apoyar la hipótesis de la profundización: la identificación de la localización de información, el etiquetado de medios gubernamentales, el bloqueo a anuncios de medios extranjeros, y la suspensión de publicidad política en el caso de Facebook, o el etiquetado de contenidos políticos,

¹⁴⁸ C. Newton, *The Coronavirus Is Forcing Tech Giants To Make A Risky Bet On Ai*, THE VERGE, 18/03/2020, disponible en <https://www.theverge.com/interface/2020/3/18/21183549/coronavirus-content-moderators-facebook-google-twitter> Fecha de consulta: 27/agosto/2020.

la suspensión de la publicidad política o el etiquetado de medios gubernamentales en el caso de Twitter muestran acciones del último año que van en esa dirección. Ellas ocurren en un marco de presión creciente sobre el modelo de negocios que suponen las redes sociales, amenazas diversas de regulación y hasta investigaciones de autoridades regulatorias por prácticas anticompetitivas.¹⁴⁹

La seguidilla de acciones de los últimos años muestran a empresas muy activas ante un fenómeno que causa preocupación y sobre el que aún se conoce poco. Ello dificulta arribar a conclusiones definitivas, pero algunas cuestiones que parecen relevantes pueden plantearse al menos a modo tentativo. Por ejemplo, ¿cuál es el impacto de estas acciones en América Latina? La respuesta usual de las plataformas—consultadas para este documento, pero también expresadas en el marco de otras iniciativas de “múltiples partes interesadas”—es que las políticas de las compañías son globales. Sin embargo, hubo muchas acciones identificadas cuya implementación en América Latina no pudo ser verificada, como surge de la Figura 4. Claramente existe cierto *gap* en los anuncios globales y su implementación en América Latina.¹⁵⁰

Por otro lado, ¿existen buenas prácticas que es posible extraer de las acciones identificadas? Ésta es una pregunta cuya respuesta también sólo puede ser tentativa. Si por *buenas prácticas* nos referimos a las acciones que son respetuosas de los estándares de derechos humanos vigentes, la respuesta podría ser positiva: claramente existen acciones más o menos problemáticas bajo ese parámetro. Pero si lo que estamos buscando son acciones efectivas, la respuesta es más esquiva porque bajo esa luz estas acciones son difíciles de evaluar: faltan estudios de impacto, no hay suficiente acceso a la información necesaria por parte de las compañías y tampoco parece haber una preocupación específica por medir resultados. Por el contrario, vistas en su conjunto las acciones parecen sucederse unas a otras en un esfuerzo por dar respuestas a presiones y malestares crecientes.

Una hipótesis que explica el volumen de acciones emprendidas se relaciona con el siguiente hecho: dentro de las múltiples presiones que reciben las compañías, la

¹⁴⁹ C. Newton, *The Antitrust Case Against Google Is Gathering Steam*, THE VERGE, 15/07/2020, disponible en <https://www.theverge.com/interface/2020/7/15/21324105/google-antitrust-california-search-youtube> Fecha de consulta: 22/septiembre/2020; A. Satariano, «Facebook Loses Antitrust Decision In Germany Over Data Collection», THE NEW YORK TIMES, 23/6/2020, disponible en <https://www.nytimes.com/2020/06/23/technology/facebook-antitrust-germany.html> Fecha de consulta: 22/septiembre/2020.

¹⁵⁰ Ver, al respecto, la respuesta de Twitter a la consulta realizada respecto de la posibilidad de reportar contenido como falso por razones electorales. Cf. Hugo Rodríguez Nicolat, ‘Comentarios de Twitter a La Encuesta de Investigación: “Desinformación En Internet: Acciones y Reacciones”’ (2020, en archivo con los autores).

amenaza de acción regulatoria está siempre presente. Al mostrar proactividad y receptividad a las demandas de actores poderosos, las compañías bajo análisis quitan argumentos a esos esfuerzos. Por otro lado, esa permeabilidad no puede sino que ser valorada por agentes estatales que encuentran—a través de la influencia informal—un canal efectivo para ejercer cierto control sobre el flujo de información en Internet. La concentración facilita ese resultado: las grandes plataformas son *de facto* intermediarios necesarios que se vuelven nodos de control influyentes.

Este escenario plantea dilemas profundos respecto del rol de los intermediarios en el flujo de información en Internet. Si esta pregunta tenía una respuesta definida en 1996 a través de las garantías de la sección 230 de la *Communications Decency Act*, resulta claro que ello ya no es así. El rol y responsabilidad de las grandes plataformas de Internet se encuentra en plena disputa en todo el mundo y algún tipo de innovación regulatoria se impondrá en un futuro que estimamos no muy lejano. Aún desconocemos, sin embargo, la forma que adquirirá esa regulación: si continuaremos con acciones de autorregulación como las que vimos hasta ahora—cuya intensidad y variedad parece haberse acelerado en el marco de la pandemia—o si una regulación *tradicional* se impondrá de la mano de estados con capacidad de incidir en lo que las plataformas hacen, como los Estados Unidos o la Unión Europea. Probablemente la cuestión será abordada con una mezcla de aproximaciones, como sugieren Marsden, Meyer y Brown: una auto-regulación controlada, una autorregulación formalizada o una co-regulación nos parecen los escenarios más factibles.¹⁵¹ En este sentido, lo que estamos viendo actualmente es—posiblemente—un lento desplazamiento de la discusión pública en esa dirección (Figura 4).

Existen señales que apoyan esa hipótesis. En efecto, si bien la base legal de las acciones de moderación continúa siendo el derecho de propiedad de las empresas y la facultad derivada de definir qué contenidos son aceptables o no, gran parte de esas decisiones se basan en requerimientos legales de países relevantes, como ocurre—por ejemplo—en los no problemáticos casos de pornografía infantil o en los más problemáticos casos de discurso “de odio”, gracias a las regulaciones vigentes especialmente en países europeos. En el menú de opciones de Marsden et al. nos encontramos actualmente entre la segunda etapa de autorregulación y la tercera, donde empiezan a incidir actores externos con p.ej., las auditorías de

¹⁵¹ C. Marsden; T. Meyer; I. Brown, «Platform Values And Democratic Elections: How Can The Law Regulate Digital Disinformation?», *COMPUTER LAW & SECURITY REVIEW*, vol. 36, 2020, Disponible en <http://www.sciencedirect.com/science/article/pii/S026736491930384X>.

GNI o el *Code of Practice on Disinformation* de la Comisión Europea.¹⁵²

Ese posible escenario futuro importará un cuestionamiento a los principios de libertad de expresión que, hasta el momento, han definido este tema. La no responsabilidad de intermediarios por los contenidos producidos por terceros operaba como un principio viable en el marco de una red descentralizada. Los procesos de concentración y centralización que caracterizaron al *giro pesimista* respecto de Internet ponen en duda que con ese principio sea suficiente para abordar múltiples problemas actuales, desde la desinformación hasta la creación algorítmica de nuevos contenidos. En todo caso, esa discusión se dará un escenario complejo, posiblemente en modelos regulatorios distintos a los modelos “puros” que imaginamos hasta el momento.¹⁵³

En este escenario los países de América Latina tendrán un rol secundario, y ello revela la persistencia de un desafío estructural en todo lo vinculado a la cuestión de la regulación de las grandes plataformas de Internet desde países periféricos: no todos los países tienen la misma capacidad de incidir en cómo estas compañías actúan.¹⁵⁴ En este sentido, resulta difícil imaginar que los países de la región puedan modificar radicalmente las grandes tendencias que se pueden observar hoy, pero quizás puedan incidir. En los últimos años hemos visto que las plataformas han invocado algunas experiencias locales como dinámicas virtuosas, como ocurrió con las colaboraciones de las plataformas con las autoridades electorales en recientes procesos electorales de México y Brasil. Hasta el momento, no hemos observado iniciativas de regulación “tradicional” que introduzcan innovaciones dignas de destacar; por el contrario, la última experiencia en ese sentido fue el proyecto de ley de Brasil sobre *Liberdade, Responsabilidade e Transparência Digital na Internet*, que fue unánimemente rechazado por la sociedad civil de la región por afectar los principios fundamentales de libertad de expresión que, por el momento, controlan el tipo de respuestas aceptables a estos problemas.

Desde este punto de vista—es decir, desde el punto de vista de lo que los estándar-

¹⁵² European Commission, «EU Code Of Practice On Disinformation». European Commission, Brussels. septiembre de 2018; C. Marsden; T. Meyer; I. Brown, “Platform values and democratic elections”, cit., pág. 12.

¹⁵³ Sobre esta cuestión, ver el modelo de Abbott y Snidel; cf. K. Abbott; D. Snidal; Walter Mattli, Ngaire Woods (eds.) («The Governance Triangle: Regulatory Standards Institutions And The Shadow Of The State», en *The politics of global regulation*, Princeton University Press, Princeton, 2009.)

¹⁵⁴ Se trata éste de un problema global, vinculado al rol que juegan las empresas transnacionales en el respeto o violación de derechos humanos en países periféricos. Al respecto, ver J. Ruggie («Protect, Respect And Remedy: A Framework For Business And Human Rights». Human Rights Council. report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, Geneve. A/HRC/8/5. 7 de abril de 2008.), pág. 14

res actuales de libertad de expresión sostienen—las diversas acciones de las plataformas encajan fácilmente dentro de un espectro imaginario de *aceptabilidad*. Las acciones comprendidas en la categoría (1) son las menos problemáticas y respetan el principio fundamental de que la mejor solución a los abusos de la libertad de expresión para por *más libertad de expresión*.¹⁵⁵ A los errores se responde con correcciones, a la información falsa se la combate con información verdadera, etcétera. Las acciones de moderación que se limitan a brindar más información, en principio, también participan de la misma conclusión normativa. El escenario empieza a cambiar, sin embargo, en casos en que esas acciones de moderación impacten en p.ej., el nivel de alcance que pueda llegar a tener el contenido señalado como—por caso—la decisión de “bajar el volumen” de contenidos problemáticos.¹⁵⁶ Ese tipo de efectos, en general difíciles de identificar, impactan en la forma en que circula la información y vuelve a los intermediarios en verdaderos “curadores” del debate público, que es el rol que—como vimos—de manera creciente están asumiendo. En esos casos, esas acciones que impactan en el debate público se vuelven más problemáticas y deben ser sometidas a los análisis normativos que exige el sistema interamericano que pasan, fundamentalmente, por utilizar criterios transparentes, utilizarlos de manera consistente y garantizando el debido proceso para los usuarios cuyos contenidos son afectados.¹⁵⁷

La transparencia sobre las acciones de moderación debería mejorar sustantivamente. Los informes de transparencia son difíciles de analizar y ofrecen la información en formatos difíciles de leer o excesivamente agregada. Como regla general, desconocemos casos específicos de moderación: para resolver ese vacío las plataformas deberían abrirse a estudios de actores independientes y académicos.¹⁵⁸ Por otro lado, las reglas de moderación también deberían ser más claras y su aplicación debería ser consistente: las plataformas aquí analizadas no son bien evaluadas por actores independientes.¹⁵⁹ Todas estas necesidades, que aquí señalamos superficialmente, se vinculan precisamente con el rol de *curación* que han adoptado las plataformas en el complejo y cambiante escenario regulatorio

¹⁵⁵ CIDH, «Marco Jurídico Interamericano del Derecho a la Libertad de Expresión». Relatoría Especial para la Libertad de Expresión de la Comisión Interamericana de Derechos Humanos. OEA/Ser.L/V/II CIDH/RELE/INF. 2/09. 30 de diciembre de 2009. Párr. 108 y ss.

¹⁵⁶ YouTube, “Continuing Our Work To Improve Recommendations On Youtube”, cit.

¹⁵⁷ CIDH, «Libertad de Expresión e Internet». Comisión Interamericana de Derechos Humanos, Washington, DC. 2013. Párr. 55; «Estándares Para Una Internet Libre, Abierta e Incluyente». Relatoría Especial para la Libertad de Expresión de la CIDH, Washington D.C. INF.17/17. 2017. Párr. 87.

¹⁵⁸ Existe una iniciativa que va en esa dirección: Social Science One. Ver <https://socialscience.one/>

¹⁵⁹ Al respecto, ver [tos;dr](https://tosdr.org/), en <https://tosdr.org/>

que señalamos en párrafos anteriores.

Aún no sabemos si ese rol va a ser aceptado por la regulación que eventualmente se imponga. Imaginamos otros mundos posibles, en los que ese rol de curación sea rechazado por innecesario: en una red menos concentrada y más descentralizada, parecida al modelo original, ese rol de curación sería inviable y poco eficiente. La información circularía de una manera un tanto más caótica y no habría formas sencillas de ejercer el control que hoy se demanda a estos actores centrales.

Ese futuro, sin embargo, parece lejano. Si bien hay investigaciones de autoridades de defensa de la competencia en todo el mundo (especialmente en Europa), vemos incentivos alineados para mantener y resguardar el rol central que los actores aquí analizados han conseguido. Para los estados resulta más sencillo controlar el flujo de información cuando hay actores centrales con capacidad de control que cuando ellos están ausentes, son demasiados o no concentran porciones significativas del tráfico. La coincidencia de esos intereses con los intereses privados de algunas de las corporaciones más poderosas del mundo sugiere que el camino regulatorio actual está dirigido a tomar a estas plataformas (y a este nivel de concentración) como un fenómeno dado. En la medida en que esta característica de Internet subsista, y en la medida en que actores diversos—organismos internacionales, ONG, centros de investigación y estados—continúen reclamando a las plataformas acciones concretas en contra de amenazas a la democracia percibidas como graves, la mayor transparencia respecto de los criterios utilizados para la moderación parece una demanda razonable. En ese contexto, las plataformas asumen un rol cada vez más parecido al de los “foros públicos” que están sometidos a criterios y estándares que no están enteramente bajo su control. Ese punto de llegada no es inexorable, y quizás no sea deseable. Pero nos parece que hacia allí vamos.