



# ¿La nueva normalidad? Desinformación y control de contenidos en las redes sociales con ocasión del covid-19

---

*Marzo de 2021*

---

**Facultad de Derecho**

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

**UP**  
**Universidad  
de Palermo**

# ¿La nueva normalidad?

## Desinformación y control de contenidos en las redes sociales con ocasión del covid-19\*

*Carlos Cortés y Luisa Isaza*

### I. Presentación

Este documento aborda las medidas que Facebook, Twitter y YouTube implementaron para enfrentar los contenidos problemáticos relacionados con el covid-19 en sus plataformas. Conocidas principalmente como normas comunitarias, estas reglas fundamentan las acciones de moderación que estos servicios ejecutan sobre el contenido de los usuarios. El objetivo principal de esta publicación es entender el impacto que tuvo el COVID-19 en las normas comunitarias.

En primer lugar, describimos cuál era el estado de las normas relevantes antes de la pandemia. En segundo lugar, explicamos qué ajustes hicieron estas plataformas para el contenido orgánico. En tercer lugar, abordamos la excepción de interés público, que protege publicaciones aun si violan las normas comunitarias. En cuarto lugar, describimos algunas medidas relacionadas con el contenido publicado en estos servicios. Finalmente, ofrecemos algunas conclusiones.

Este análisis no constituye un insumo exhaustivo del tema ni tampoco abarca todas las modificaciones que posiblemente se registraron en esta materia. Precisamente, uno de los problemas identificados a lo largo de esta investigación es la dificultad para entender dónde están, cómo cambian y de qué manera se implementan las normas comunitarias de las redes sociales. Por otro lado, este documento no incluye temas como *deep-fakes*, contenido multimedia manipulado u operaciones de influencia.

La observación se hizo con corte al 31 de diciembre de 2020 y se sirvió del monitoreo de las normas comunitarias realizado por el CELE y Linterna Verde a través del proyecto Letra Chica.<sup>1</sup> Letra Chica hace seguimiento, explica y pone en contexto los cambios a las normas comunitarias de Facebook, Twitter y YouTube.<sup>2</sup> Las medidas que abarca el texto se resumen también en tablas incluidas como anexos.

<sup>1</sup> Linterna Verde es un grupo interdisciplinario sin ánimo de lucro que resuelve preguntas sobre el debate público digital.

<sup>2</sup> Para más información, ver: <https://letrachica.digital>, último acceso: 9 de marzo de 2021.

Este documento se finalizó mientras sucedió el hecho histórico de que varias redes sociales sancionaron las cuentas del entonces presidente de Estados Unidos, Donald Trump. En particular, Twitter suspendió su cuenta de manera definitiva mientras que Facebook remitió esa misma decisión para revisión por parte de la junta de supervisión que entró en funcionamiento el año pasado.<sup>3</sup> Es importante aclarar, entonces, que esa coyuntura no está comprendida en este texto. No obstante, la crisis del covid-19 coincidió con la campaña presidencial en Estados Unidos, y en esa medida algunos elementos relacionados con ese debate sí están reflejados en este estudio.

## II. El mundo antes de la pandemia

Antes de que el nuevo coronavirus fuera declarado como pandemia por la Organización Mundial de la Salud (OMS) en marzo de 2020, la desinformación ya era un problema lo suficientemente complejo para las plataformas de internet. Un mes antes, cuando la crisis apenas comenzaba, el director de la OMS, Tedros Adhanom Ghebreyesus, lo advirtió: “Esta infodemia está obstaculizando las medidas de contención del brote, propagando pánico y confusión de forma innecesaria y generando división”.<sup>4</sup>

Los elementos que señalaba el director de la OMS bien podrían aplicarse a lo que venía pasando antes de esta emergencia de salud pública: la desinformación es un obstáculo para la contención de la pandemia, sirve para difundir pánico y divide a la ciudadanía.<sup>5</sup> Sin duda, el gran punto de tensión previo al covid-19 fueron las elecciones presidenciales estadounidenses de 2016, donde la campaña de Donald Trump cabalgó sobre una estrategia de manipulación cuyo impacto en el resultado sigue siendo objeto de estudio.<sup>6</sup>

<sup>3</sup> Ver, Clegg, Nick, “Referring Former President Trump’s Suspension from Facebook to the Oversight Board”, Facebook, 21 de enero, 2021, disponible en: <https://about.fb.com/news/2021/01/referring-trump-suspension-to-oversight-board>, último acceso: 4 de marzo de 2021.

<sup>4</sup> Adhanom Ghebreyesus, Tedros y NG, Alex, “Desinformación frente a medicina: hagamos frente a la ‘infodemia’”, *El País*, 18 de febrero, 2020, disponible en: [https://elpais.com/sociedad/2020/02/18/actualidad/1582053544\\_191857.html](https://elpais.com/sociedad/2020/02/18/actualidad/1582053544_191857.html), último acceso: 4 de marzo de 2021.

<sup>5</sup> Sobre las medidas tomadas por Facebook, Google y Twitter para combatir la desinformación antes de la pandemia, ver: Cortés, Carlos y Isaza, Luisa, *Noticias falsas en Internet: la estrategia para combatir la desinformación*, CELE, diciembre, 2017, disponible en: <https://www.palermo.edu/cele/pdf/FakeNews.pdf>, último acceso: 4 de marzo de 2021.

<sup>6</sup> Algunos estudios con conclusiones diversas sobre la influencia de las noticias falsas en la elección: Gunther, Richard, Nisbet, Erik C. y Beck, Paul, “Trump May Owe his 2016 Victory to ‘Fake News’, New Study Suggests”, *The Conversation*, 15 de febrero, 2018, disponible en: <https://theconversation.com/trump-may-owe-his-2016-victory-to-fake-news-new-study-suggests-91538>, último acceso: 4 de marzo de 2021. Guess, Andrew M., Nyhan, Brendan y Reifler, Jason, “Exposure to Untrustworthy Websites in the 2016 US Election”, *Nature Human Behaviour*, 2 de marzo, 2020, disponible en: <https://www.nature.com/articles/>

Por supuesto, las presiones de los Estados a las plataformas aumentaron. Según académicos de la Universidad de Oxford, en los dos años siguientes a las elecciones de 2016 de los Estados Unidos, por lo menos 43 países alrededor del mundo propusieron o implementaron regulaciones específicamente pensadas para controlar las operaciones de influencia y la desinformación.<sup>7</sup>

Pero las críticas a las redes sociales –y en especial a Facebook– por la desinformación que se difunde a través de ellas no se limita a lo acontecido en las elecciones estadounidenses. En Birmania (Myanmar), el acceso tardío e intempestivo a internet (en 2014, el 1% de la población usaba internet; en 2016, el 20%),<sup>8</sup> sumado a una estrategia de propaganda anti musulmana en redes sociales, desembocó en estigmatizaciones, ataques y violencia contra los musulmanes rohinyá.<sup>9</sup> En agosto de 2018, tras el asesinato de más de 25.000 rohinyá y el desplazamiento forzado de otros 700.000, Facebook reconoció que se había demorado en actuar ante la desinformación y el odio antimusulmán en este país.<sup>10</sup> Como consecuencia, suspendió tardíamente decenas de usuarios y páginas vinculadas al ejército birmano –muchas con millones de seguidores– por violar las políticas sobre integridad y autenticidad en la plataforma.<sup>11</sup>

En América Latina también se ha llamado la atención sobre la desinformación y la actividad inauténtica en redes sociales en el marco de contextos electorales. El

---

s41562-020-0833-x?proof=trueMay%252F, último acceso: 4 de marzo de 2021. Guess, Andrew, Nagler, Jonathan y Tucker, Joshua, “Less than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook”, *Science Advances*, 9 de enero, 2019, disponible en: <https://advances.sciencemag.org/content/5/1/eaau4586>, último acceso: 4 de marzo de 2021

<sup>7</sup> Bradshaw, Samantha, Neudert, Lisa-María y Howard, Philip N., *Government Responses to Malicious Use of Social Media*, Oxford Internet Institute, University of Oxford, 2019, disponible en: <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/01/Nato-Report.pdf>, último acceso: 4 de marzo de 2021. El estudio se centró en los cien países con mayores usuarios de internet en 2016, dentro de los cuales identificaron los 43 países mencionados. Dentro de la regulación dirigida a redes sociales se encontraron medidas para la remoción de contenidos, la transparencia en la publicidad política y la protección de datos personales.

<sup>8</sup> Frenkel, Sheera, “This Is What Happens When Millions of People Suddenly Get the Internet”, *Buzz Feed News*, 20 de noviembre, 2016, disponible en: <https://www.buzzfeednews.com/article/sheerafrenkel/fake-news-spreads-trump-around-the-world>, último acceso: 4 de marzo de 2021.

<sup>9</sup> *Ibid.* Ver también: Gowen, Annie y Bearak, Max, “Fake News on Facebook Fans the Flames of Hate against the Rohingya in Burma”, *The Washington Post*, 8 de diciembre, 2017, disponible en: [https://www.washingtonpost.com/world/asia\\_pacific/fake-news-on-facebook-fans-the-flames-of-hate-against-the-rohingya-in-burma/2017/12/07/2c1fe830-ca1f-11e7-b506-8a10ed11ecf5\\_story.html](https://www.washingtonpost.com/world/asia_pacific/fake-news-on-facebook-fans-the-flames-of-hate-against-the-rohingya-in-burma/2017/12/07/2c1fe830-ca1f-11e7-b506-8a10ed11ecf5_story.html), último acceso: 4 de marzo de 2021.

<sup>10</sup> Facebook, “Removing Myanmar Military Officials from Facebook”, 28 de agosto, 2018, disponible en: <https://about.fb.com/news/2018/08/removing-myanmar-officials>, último acceso: 4 de marzo de 2021.

<sup>11</sup> En agosto de 2020, Facebook anunció cómo se está preparando para combatir el discurso de odio y la desinformación antes de las elecciones de Birmania de noviembre de 2020, las segundas elecciones democráticas de la historia reciente de este país. Frankel, Rafael, “How Facebook Is Preparing for Myanmar’s 2020 Election”, Facebook, 31 de agosto, 2020, disponible en: <https://about.fb.com/news/2020/08/preparing-for-myanmars-2020-election>, último acceso: 4 de marzo de 2021.

centro de investigación InternetLab analizó los perfiles de los seguidores en Twitter de los candidatos presidenciales en Brasil en 2018 para detectar qué porcentaje de ellos correspondía a cuentas potencialmente automatizadas: el candidato con el menor número de bots tenía 13%, mientras que el que más tenía alcanzaba 63%.<sup>12</sup> A la par con esto, son usuales las operaciones de influencia en redes sociales en países como México, Colombia y El Salvador, entre otros, donde cuentas tipo “call center” capturan etiquetas y tendencias y atacan coordinadamente a críticos, periodistas y líderes políticos.

A pesar de las denuncias y presiones de distintos sectores, las plataformas venían asumiendo un rol más bien pasivo frente a los contenidos desinformativos, al centrar su intervención en la autenticidad de las cuentas y la visibilidad de las publicaciones. Los líderes de estas empresas plantearon, no sin razón, que estas plataformas debían evitar convertirse en jueces del debate público.

Twitter fue la que más resistió ese rol: “Lo que podríamos hacer es ayudar a proporcionar más contexto, ya sea mostrando todas las diferentes perspectivas (...). Pero pienso que sería peligroso para una empresa como la nuestra ser árbitros de la verdad”, dijo el fundador y director de Twitter, Jack Dorsey, en agosto de 2018.<sup>13</sup> En esa misma línea, Mark Zuckerberg, fundador y director ejecutivo de Facebook, sostuvo –en un conocido discurso dado en la Universidad de Georgetown en octubre de 2019– que aunque le preocupaba la erosión de la verdad, no creía que la gente quisiera “vivir en un mundo en el que solo se pueden publicar cosas que las empresas de tecnología consideran 100% verdaderas”.<sup>14</sup>

El eje del debate para YouTube ha sido otro: a la desinformación propia de los contenidos se añade el sistema de recomendaciones de la plataforma, que genera incentivos para que se produzcan y consuman publicaciones similares, con lo cual la experiencia del usuario termina dominada por un “efecto madriguera” o *rabbit hole*.<sup>15</sup> En diciembre de 2019, Susan Wojcicki, directora ejecutiva de You-

<sup>12</sup> Lago, Lucas y Massaro, Heloisa, “Bot or not: Who Are the Followers of our Candidates for President?”, InternetLab, disponible en: <https://www.internetlab.org.br/en/information-politics/bot-or-not-who-are-the-followers-of-our-candidates-for-president>, último acceso: 4 de marzo de 2021.

<sup>13</sup> “I think what we could do is help provide more context, whether it be showing all the different perspectives (...). We have not figured this out, but I do think it would be dangerous for a company like ours... to be arbiters of the truth”. Traducción informal. Entrevista con CNN, agosto, 2018, disponible en: [https://youtu.be/Cm\\_lmWWKDug?t=503](https://youtu.be/Cm_lmWWKDug?t=503), último acceso: 4 de marzo de 2021.

<sup>14</sup> Traducción informal. “While I certainly worry about an erosion of truth, I don’t think most people want to live in a world where you can only post things that tech companies judge to be 100 percent true”. Discurso en la Universidad de Georgetown, octubre, 2019, disponible en: <https://youtu.be/2MTpd7YOnyU?t=1802>, último acceso: 4 de marzo de 2021.

<sup>15</sup> Fox, Chris, “YouTube: ‘We Don’t Take You down the Rabbit Hole’”, BBC News, 19 de julio, 2019, disponible en: <https://www.bbc.com/news/technology-49038155>, último acceso: 4 de marzo de 2021.

Tube, dijo sobre este punto que una intervención fuerte de la plataforma podría perjudicar la disponibilidad de contenidos relevantes para los usuarios: “Si nosotros fuéramos responsables por cada pieza de contenido que recomendamos, tendríamos que revisarlas. Eso significaría que habría un número mucho menor de información que la gente encontraría. Mucho, mucho menor”.<sup>16</sup>

La posición de los líderes de estas empresas se explica en lógicas comerciales y políticas: arbitrar los contenidos ya era de por sí una tarea costosa y dispendiosa, y entrar a evaluar la veracidad de la información terminaría enfrentando a las redes sociales con partidos políticos, Gobiernos y la sociedad civil. No obstante, también había razones de fondo sobre libertad de expresión, su debido proceso y transparencia.

En cualquier caso, en el mundo anterior a la pandemia las redes sociales se aproximaron a la desinformación en punta de pies. En términos generales, su principio había sido el de la mínima intervención en el contenido, ya fuera por no tener normas sobre el tema o por su aplicación inconsistente. En palabras del periodista Casey Newton, “cada plataforma tecnológica tiene dos políticas sobre lo que permiten: la política que está por escrito y la política que es aplicada. Idealmente no habría brecha entre ellas, pero en la práctica esto casi no se puede evitar”.<sup>17</sup>

Frente a la emergencia del coronavirus, las redes sociales cambiaron su enfoque sobre la marcha, al diseñar e implementar medidas que hasta hace poco parecían imposibles. Para hacerlo, echaron mano de dos tipos de reglas: las que se enfocan en identificar acciones inauténticas en la plataforma y las que apuntan a juzgar el contenido. Basándose en sus normas comunitarias, las plataformas cuentan con procesos de moderación de contenidos que dependen en gran medida del análisis humano. Se trata de auténticos ejércitos de personas entrenadas para tomar decisiones complejas en escala masiva, sin contexto suficiente y bajo mucha presión emocional. Y si esta tarea era difícil cuando el foco eran asuntos teóricamente más sencillos –como suplantación de identidad, divulgación de información privada o amenazas–, la desinformación alrededor de la pandemia simplemente puso de presente lo inabarcable del problema.

---

<sup>16</sup> Traducción informal. “If we were held liable for every single piece of content that we recommended, we would have to review it. That would mean there'd be a much smaller set of information that people would be finding. Much, much smaller”. Entrevista con “60 Minutes”. Loizos, Connie, “In ‘60 Minutes’ Appearance, YouTube’s CEO Offers a Master Class in Moral Equivalency”, Tech Crunch, 1º de diciembre, 2019, disponible en: <https://techcrunch.com/2019/12/01/in-60-minutes-appearance-youtubes-ceo-offers-a-master-class-in-moral-equivalency>, último acceso: 4 de marzo de 2021.

<sup>17</sup> Newton, Casey, “Getting Rid of QAnon Won’t Be as Easy as Twitter Might Think”, The Verge, 23 de julio, 2020, disponible en: <https://www.theverge.com/interface/2020/7/23/21334255/twitter-qanon-ban-facebook-policy-enforcement-political-candidates>, último acceso: 4 de marzo de 2021.

La automatización tampoco ofrece, al menos por ahora, una solución. Si bien los algoritmos pueden detectar comportamientos sospechosos, spam y ciertos tipos de contenidos –sobre todo alrededor de vídeos y fotos– no saben cómo resolver dilemas propios del contexto de una expresión, ni mucho menos decidir sobre la veracidad de una información. El riesgo latente es generar un alto porcentaje de “falsos positivos”, con los consecuentes efectos negativos en términos de inhibición y censura.

En esencia, el covid-19 apareció mientras las redes sociales apagaban incendios, dándole prioridad a los mercados más relevantes y a los escándalos noticiosos más dañinos. Este modo de control de daños ha estado marcado por reglas y procesos inconsistentes cuya transparencia avanza de manera lenta. A continuación, veremos cuáles eran las reglas de Facebook, Twitter y YouTube para enfrentar la desinformación antes de que llegara la pandemia.

## 1. Facebook

Para Facebook, en principio la publicación de noticias falsas no viola sus normas comunitarias.<sup>18</sup> Por lo tanto, no elimina este contenido, sino que se reduce su distribución.<sup>19</sup> Y para hacerlo, la compañía se basa principalmente en la verificación adelantada por terceros.

En un proceso conocido como *fact-checking*, organizaciones de la sociedad civil en todo el mundo evalúan miles de publicaciones, principalmente de medios de comunicación.<sup>20</sup> Con ese insumo, Facebook implementa alguna acción que afecta la visibilidad y la distribución del contenido evaluado: alerta a la persona que lo ve o que lo va a compartir; “castiga” de manera parcial o total su visibilidad en el sector de las noticias, e incluso puede sancionar las cuentas que lo crean o compartan repetidamente.<sup>21</sup> Estas medidas, sin embargo, no son consistentes en todo el mundo.

---

<sup>18</sup> “False news does not violate our Community Standards”. Lyons, Tessa, “Hard Questions: What’s Facebook’s Strategy for Stopping False News?”, Facebook, 23 de mayo, 2018, disponible en: <https://about.fb.com/news/2018/05/hard-questions-false-news>, último acceso: 4 de marzo de 2021.

<sup>19</sup> Facebook, “21. Noticias falsas”, disponible en: [https://www.facebook.com/communitystandards/false\\_news](https://www.facebook.com/communitystandards/false_news), último acceso: 4 de marzo de 2021.

<sup>20</sup> Más información sobre el programa de *fact-checking* y la estrategia de Facebook para reducir las noticias falsas: Facebook, “Verificación de datos en Facebook”, disponible en: <https://www.facebook.com/business/help/2593586717571940>, último acceso: 4 de marzo de 2021, y Lyons, “Hard Questions: What’s Facebook’s Strategy for Stopping False News?”, *op. cit.*

<sup>21</sup> Para obtener información más detallada sobre las medidas adoptadas por Facebook para controlar la desinformación inmediatamente después de las elecciones estadounidenses de 2016, ver: Cortés e Isaza, *Noticias falsas en Internet: la estrategia para combatir la desinformación*, *op. cit.*



Hay otras normas y anuncios que son relevantes:

- Facebook tiene un grupo de reglas sobre “organización de actos para infringir daños y publicidad de la delincuencia”.<sup>22</sup> Allí se prohíbe la divulgación de contenidos que muestren, confiesen o fomenten actos de daños físicos hacia seres humanos, llamadas falsas a servicios de emergencia o participación en desafíos virales de alto riesgo.<sup>23</sup>
- En el capítulo de “Artículos regulados”, Facebook prohíbe la publicación de contenidos que busquen comprar, vender, intercambiar, donar, regalar, solicitar u ofrecer fármacos.
- En una entrada de blog –es decir, por fuera de las normas comunitarias– Facebook informó que desde 2018 se borra la desinformación sobre salud que pueda contribuir a un daño físico inminente.<sup>24</sup> En esta misma entrada, posteriormente aclaró que desde enero de 2020 ha borrado publicaciones falsas sobre el covid-19. A este punto volveremos más adelante.

Más allá de estos elementos, Facebook ha apuntado a controlar la desinformación al enfocarse en identificar las “operaciones de información o influencia”, o sea, las acciones coordinadas que se valen de cuentas automatizadas y humanas para amplificar contenidos, intimidar a otros usuarios y capturar conversaciones y tendencias. Cabe recordar que esta discusión se enmarcó en serias acusaciones de interferencia rusa en las elecciones de Estados Unidos de 2016.

En octubre de 2019, Mark Zuckerberg afirmó que esta era “una solución mucho mejor que la definición en constante expansión de qué constituye un discurso dañino”.<sup>25</sup> Siguiendo esa línea, para Zuckerberg el verdadero problema de las publica-

---

<sup>22</sup> Facebook, “2. Personas y organizaciones peligrosas”, disponible en: [https://www.facebook.com/communitystandards/dangerous\\_individuals\\_organizations](https://www.facebook.com/communitystandards/dangerous_individuals_organizations), último acceso: 4 de marzo de 2021.

<sup>23</sup> Bajo esta política, Facebook prohíbe, por ejemplo, la incitación a hacer llamadas falsas a los servicios de emergencia (actividad conocida como *swatting*) o a participar en desafíos virales de alto riesgo.

<sup>24</sup> Clegg, Nick, “Combating Covid-19 Misinformation across our Apps”, Facebook, 25 de mayo, 2020, disponible en: <https://about.fb.com/news/2020/03/combating-covid-19-misinformation>, último acceso: 5 de marzo de 2021. Según Facebook, desde 2018 ha borrado, entre otros, contenidos sobre el sarampión en Samoa y la vacuna contra el polio en Pakistán.

<sup>25</sup> Romm, Tony, “Zuckerberg: Standing for Voice and Free Expression”, *The Washington Post*, 17 de octubre, 2019, disponible en: <https://www.washingtonpost.com/technology/2019/10/17/zuckerberg-standing-voice-free-expression>, último acceso: 5 de marzo de 2021. Sarah C. Haan presenta una posición bastante crítica de esta estrategia en: Haan, Sarah C., “The Authenticity Trap. Mark Zuckerberg Thinks Facebook’s Problems Can Be Fixed with ‘Authentic’ Speech. He’s so Wrong”, *Slate*, 21 de octubre, 2019, disponible en: <https://slate.com/technology/2019/10/mark-zuckerberg-facebook-georgetown-speech-authentic.html>, último acceso: 5 de marzo de 2021.



ciones rusas que buscaron interferir en las elecciones no fue propiamente su contenido, sino el hecho de que fueran hechas de manera coordinada por cuentas falsas.

Para comienzos de 2020, Facebook ya contaba entonces con reglas sobre representaciones engañosas y comportamientos inauténticos.<sup>26</sup> Primero –y a diferencia de Twitter– Facebook tiene una política de nombres reales según la cual, primero, los usuarios deben utilizar la identidad con la que se reconocen de manera cotidiana; segundo, la plataforma no permite que sus usuarios publiquen, interactúen con contenido o creen cuentas con una frecuencia alta; tercero, hay reglas anti-spam: los usuarios no pueden requerir o engañar a otros para que interactúen con ciertos contenidos. Finalmente, Facebook prohíbe participar en comportamientos no auténticos, individuales o coordinados, para engañar a las personas respecto de la popularidad o el origen de un contenido.<sup>27</sup>

## 2. Twitter

Siguiendo la línea planteada por Jack Dorsey según la cual Twitter no debía convertirse en árbitro de la verdad, la plataforma quiso enfrentar la desinformación al enfocarse en la actividad de las cuentas. Es decir, se centró en los actores antes que en el contenido.

Durante la campaña presidencial de 2016, el uso de bots en Twitter no se dio de manera aislada. En junio de 2017, cuando se discutía la responsabilidad de las redes sociales en la interferencia rusa del año anterior en Estados Unidos, Twitter reivindicó el carácter abierto de su plataforma y anunció que concentraría sus esfuerzos en evitar la automatización con fines de manipulación: “Prohibimos estrictamente el uso de bots y otras redes de manipulación para socavar la funcionalidad principal de nuestro servicio”, afirmó la compañía en su blog oficial.<sup>28</sup> En ese contexto, para 2018 el objetivo de Twitter era detectar esas operaciones para evitar la desinformación, los ataques y el spam.<sup>29</sup>

<sup>26</sup> Facebook, “17. Integridad de la cuenta y autenticidad de identidad”, disponible en: [https://www.facebook.com/communitystandards/integrity\\_authenticity](https://www.facebook.com/communitystandards/integrity_authenticity), último acceso: 5 de marzo de 2021.

<sup>27</sup> Facebook, “20. Comportamiento no auténtico”, disponible en: [https://www.facebook.com/communitystandards/inauthentic\\_behavior](https://www.facebook.com/communitystandards/inauthentic_behavior), último acceso: 20 de marzo de 2021.

<sup>28</sup> Crowell, Colin, “Our Approach to Bots and Misinformation”, Twitter, 14 de junio, 2017, disponible en: [https://blog.twitter.com/official/en\\_us/topics/company/2017/Our-Approach-Bots-Misinformation.html](https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html), último acceso: 5 de marzo de 2021.

<sup>29</sup> Twitter Public Policy (“Políticas públicas de Twitter”), 30 de julio, 2020, disponible en: [https://about.twitter.com/en\\_us/advocacy/elections-integrity.html#us-elections](https://about.twitter.com/en_us/advocacy/elections-integrity.html#us-elections), último acceso: 5 de marzo de 2021. Hoy en día, Twitter aborda este problema en informes sobre manipulación de la plataforma. El último disponible es del segundo semestre de 2019: Twitter, “Platform Ma-

Sin embargo, antes de 2020 Twitter sí empezó a incluir algunas prohibiciones relacionadas con la desinformación, específicamente en el contexto de procesos electorales. En 2019, Twitter creó una política sobre “integridad de las elecciones” en la que se prohibía la publicación de información engañosa sobre cómo participar, contenidos de supresión e intimidación de votantes e información falsa o engañosa sobre la afiliación política.<sup>30</sup> Ese mismo año, con ocasión de las elecciones en la Unión Europea, la plataforma introdujo la posibilidad de que los usuarios reportaran tuits infractores de esta política. Como ha sido una constante entre estas empresas, el cambio tenía un enfoque geográfico y no implicaba un ajuste en las normas comunitarias. Para 2020, con el censo estadounidense y la campaña electoral presidencial en el horizonte, Twitter amplió las reglas y las renombró “política de integridad cívica”.

### 3. YouTube

La estrategia de YouTube para controlar la desinformación se basa en tres principios: i) la conservación del contenido en la plataforma a menos que viole sus normas comunitarias; ii) la posibilidad de monetizar publicaciones es un privilegio; y iii) los videos deben cumplir estándares exigentes para que la plataforma los recomiende.<sup>31</sup> En la aplicación del primero de estos principios, para el control de la desinformación sobre el coronavirus varias normas comunitarias resultarían relevantes:

- En su política sobre contenido perjudicial o peligroso, YouTube prohíbe las publicaciones en las que se promoció, recomiende o afirme que el uso de sustancias o tratamientos dañinos puede ser beneficioso para la salud.<sup>32</sup>
- En su política sobre prácticas engañosas, la plataforma prohíbe el contenido manipulado o modificado que busque engañar al usuario y que pueda involucrar un grave riesgo de daño flagrante.<sup>33</sup>

---

nipulation”, 2019, disponible en: <https://transparency.twitter.com/en/reports/platform-manipulation.html#2019-jul-dec>, último acceso: 5 de marzo de 2021.

<sup>30</sup> Comisión Europea, “Fourth Intermediate Results of the EU Code of Practice against Disinformation”, 17 de mayo, 2019, disponible en: <https://ec.europa.eu/digital-single-market/en/news/fourth-intermediate-results-eu-code-practice-against-disinformation>, último acceso: 5 de marzo de 2021. Ver el reporte de abril de 2019 de Twitter, “Platform Manipulation”, *op. cit.*

<sup>31</sup> Google, “How Google Fights Disinformation”, febrero, 2019, disponible en: [https://www.blog.google/documents/37/How\\_Google\\_Fights\\_Disinformation.pdf?hl=en](https://www.blog.google/documents/37/How_Google_Fights_Disinformation.pdf?hl=en), último acceso: 5 de marzo de 2021.

<sup>32</sup> Google, “Política de contenido perjudicial o peligroso”, disponible en: [https://support.google.com/youtube/answer/2801964?hl=es-419&ref\\_topic=9282436](https://support.google.com/youtube/answer/2801964?hl=es-419&ref_topic=9282436), último acceso: 5 de marzo de 2021.

<sup>33</sup> Google, “Políticas sobre spam, prácticas engañosas y estafas”, disponible en: [https://support.google.com/youtube/answer/2801973?hl=es-419&ref\\_topic=9282365](https://support.google.com/youtube/answer/2801973?hl=es-419&ref_topic=9282365), último acceso: 5 de marzo de 2021.

- YouTube cuenta con un grupo más amplio de reglas para asegurar el comportamiento auténtico. A lo largo de diversas políticas,<sup>34</sup> YouTube prohíbe varios tipos de comportamiento que busquen engañosamente redirigir a los usuarios a otros sitios o aumentar artificialmente las métricas de participación (vistas, comentarios, “me gusta”), así como crear listas de reproducción con títulos o descripciones engañosas que hagan creer a los usuarios que verán videos diferentes a los que contiene la lista.

Para asegurar mayor calidad en los contenidos y combatir el spam y otras acciones engañosas, en aplicación del segundo principio, YouTube postula que solo los videos de calidad pueden ser monetizados.<sup>35</sup> Es decir, los creadores que pretendan monetizar sus videos no solo deben cumplir las normas comunitarias generales, sino que también deben observar reglas más estrictas sobre contenido apto para anunciantes. Una última estrategia de YouTube para mejorar la calidad de los videos que consumen los usuarios es enriquecer el sistema de recomendaciones. A partir del historial de los usuarios y otros elementos del algoritmo, la plataforma hace sugerencias personalizadas en la página principal, en los resultados de búsqueda y en el espacio “A continuación” que sale cuando está finalizando una reproducción.<sup>36</sup>

### III. Aumenta la presión, aumentan las medidas

El covid-19 se convirtió en la tormenta perfecta para las plataformas. La creciente presión que venían enfrentando para remover contenidos problemáticos se volvió un asunto de salud pública. La desinformación alrededor de la pandemia se propagaba como el virus –de la mano de líderes políticos e influenciadores–,

---

<sup>34</sup> *Ibid.* Google, “Políticade participación falsa”, disponible en: <https://support.google.com/youtube/answer/3399767?hl=es-419>, último acceso: 5 de marzo de 2021. Google, “Política sobre las listas de reproducción”, disponible en: [https://support.google.com/youtube/answer/9713446?hl=es-419&ref\\_topic=9282365](https://support.google.com/youtube/answer/9713446?hl=es-419&ref_topic=9282365), último acceso: 5 de marzo de 2021.

<sup>35</sup> Para obtener dinero con la publicidad que la plataforma muestra en sus videos, los creadores deben formar parte del programa de socios de YouTube. Para pertenecer al programa se requieren por lo menos mil suscriptores y cuatro mil horas de reproducción. Google, “Descripción general y elegibilidad del programa de socios de YouTube”, disponible en: <https://support.google.com/youtube/answer/72851?hl=es-419>, último acceso: 5 de marzo de 2021.

<sup>36</sup> Inicialmente, el sistema de recomendaciones sugería los contenidos que lograban más clics por parte de los usuarios. En consecuencia, YouTube cambió el enfoque: para identificar el contenido que más les gusta a los usuarios, se fija en el tiempo que dedican a ver el video y si ven hasta el final. Por otro lado, YouTube también busca detectar y dar prevalencia a fuentes confiables, por ejemplo, tomando la cantidad de links que apuntan a un contenido como un indicador de autoridad. Finalmente, YouTube también realiza encuestas para estudiar si los usuarios se sienten satisfechos con las recomendaciones. Google, “How Google Fights Disinformation”, *op. cit.*

las normas comunitarias de las redes sociales eran insuficientes e inconsistentes y los encargados de aplicarlas tenían que confinarse en sus casas. En medio de la confusión, rápidamente las plataformas empezaron a anunciar nuevas normas y medidas para lidiar con la desinformación. En lo que se relaciona con las normas comunitarias, las acciones comunicadas durante la pandemia se han centrado más en las reglas sobre el contenido de las publicaciones, que en las reglas sobre actividades inauténticas.<sup>37</sup> A continuación se hace un recuento de las acciones que están tomando las plataformas.<sup>38</sup> En un acápite posterior se hace un recuento de las nuevas reglas de publicidad que han creado las empresas.

## 1. Facebook

En enero de 2020, cuando el coronavirus no era aún una pandemia global, Facebook fue la primera red social en hacer anuncios sobre el control de la desinformación. En ese momento, según explicó, la plataforma usó reglas que ya tenía en marcha. En esencia, la intervención giraba alrededor del etiquetado, los filtros y la remoción de contenido:

- Facebook mantiene su trabajo con *fact-checkers* externos.<sup>39</sup> A partir del diagnóstico de los verificadores, Facebook etiqueta información falsa y limita la difusión en la plataforma. Una vez que un contenido es señalado, Facebook activa métodos de detección proactiva de posibles duplicados.<sup>40</sup>
- Siguiendo una práctica implementada desde antes de la pandemia, la plataforma muestra alertas a las personas que han compartido o que están tratando de compartir un contenido señalado como falso.
- Facebook se comprometió a eliminar contenidos falsos y teorías de la cons-

<sup>37</sup> En publicaciones en su blog en enero y marzo de 2020, Twitter sostuvo que en ese momento no estaba viendo esfuerzos significativos de manipulación coordinada de la plataforma en torno al coronavirus. Twitter, “Our Zero-tolerance Approach to Platform Manipulation”, 4 de marzo, 2020, disponible en: [https://blog.twitter.com/en\\_us/topics/company/2020/covid-19.html#zero-tolerance](https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#zero-tolerance), último acceso: 11 de marzo de 2021.

<sup>38</sup> Para el detalle de los cambios ver: CELE, Letra Chica, “Cambios por covid-19”, actualización del 12 de febrero, 2021, disponible en: <https://letrachica.digital/wiki/cambios-covid>, último acceso: 5 de marzo de 2021.

<sup>39</sup> “Nuestra red global de verificadores de datos externos continúa su trabajo revisando el contenido y desacreditando las afirmaciones falsas que se están extendiendo relacionadas con el coronavirus”, explicó la compañía en el blog sobre el tema. Jin, Kang-Xing, “Keeping People Safe and Informed about the Coronavirus”, Facebook, 18 de diciembre, 2020, “Limiting Misinformation and Harmful Content”, 30 de enero, 2020, disponible en: <https://about.fb.com/news/2020/08/coronavirus>, último acceso: 5 de marzo de 2021.

<sup>40</sup> Rosen, Guy, “An Update on our Work to Keep People Informed and Limit Misinformation about Covid-19”, Facebook, 16 de abril, 2020, disponible en: <https://about.fb.com/news/2020/04/covid-19-misinfo-update>, último acceso: 5 de marzo de 2021.

piración que pudieran causarles daño a las personas que los creyeran y que contradijeran a las autoridades de salud. El enfoque, explicó, son los contenidos que desincentivan un tratamiento adecuado o la implementación de medidas de prevención.

En su blog, Facebook informó que, siguiendo las guías de la OMS y otras autoridades de salud, desde enero ha borrado información falsa sobre curas, tratamientos, disponibilidad de servicios esenciales, la ubicación y la gravedad del brote.<sup>41</sup> En esa publicación, la plataforma explicó que esta no es una práctica nueva: desde 2018 ha eliminado la desinformación que pueda ocasionar daño físico inminente, como información falsa sobre el sarampión en Samoa o rumores sobre la vacuna contra el polio en Pakistán.

Según la compañía, estas acciones de eliminación se hacen como extensión de una regla ya existente sobre contenidos que causan daño. Y si bien esa interpretación podría hacerse a partir de esa norma, esto no se colige de manera directa ni es necesariamente obvio.<sup>42</sup>

Un mes después del primer comunicado, las medidas empezaron a incluir otros frentes, al tener en el horizonte los problemas de contenido sobre covid-19. Por ejemplo, Facebook anunció que deshabilitaría la opción de buscar efectos de realidad aumentada relacionados con el virus en Instagram; excluyó de las recomendaciones contenidos o cuentas orgánicas relacionadas con el covid-19; y, como si se tratara de una notificación de exposición al virus, Facebook informó que alertaría a las personas que hubieran interactuado (con “me gusta”, reacciones o comentarios) con un contenido que hubiese sido desacreditado –incluyendo allí recomendaciones de información confiable–. En la práctica, esta aproximación le permitió a Facebook crear una serie de reglas informales sobre la marcha para enfrentar la desinformación durante la pandemia, las cuales también hizo extensivas a Instagram.<sup>43</sup>

<sup>41</sup> Clegg, Nick, “Combating Covid-19 Misinformation across our Apps”, *op. cit.*

<sup>42</sup> “Política sobre organización de actos para infringir daños y publicidad de la delincuencia: daños a personas. Mostrar, confesar o fomentar los siguientes actos cometidos por ti o personas relacionadas contigo: actos de daños físicos hacia seres humanos, incluidos actos de violencia doméstica, excepto cuando se comparte en un contexto de redención o defensa de uno mismo u otra persona. Declarar la intención de realizar, incitar, representar, apoyar o defender, o mostrar, confesar o fomentar los siguientes actos cometidos por ti o personas relacionadas contigo: llamadas falsas a los servicios de emergencia (*swatting*). Mostrar, fomentar, defender o incitar: la participación en desafíos virales de alto riesgo”. Facebook, “3. Organización de actos para infringir daños y promoción de la delincuencia”, disponible en: [https://www.facebook.com/communitystandards/COORDINATING\\_HARM\\_PUBLICIZING\\_CRIME](https://www.facebook.com/communitystandards/COORDINATING_HARM_PUBLICIZING_CRIME), último acceso: 20 de marzo de 2021

<sup>43</sup> Facebook ha anunciado medidas similares para Instagram, uno de sus productos: i) se bloquea o restringe el uso de hashtags para esparcir desinformación; ii) se eliminan de la sección de recomendados las cuentas o contenidos relacionados con el covid-19, a menos que provengan de organizaciones de salud confiables; iii) se deshabilita la opción de buscar efectos

Hacia el final del año, en vista de la aprobación de las vacunas contra el covid-19, Facebook anunció que eliminaría las afirmaciones falsas sobre las vacunas contra el covid-19 que hayan sido desvirtuadas por expertos de salud pública.

## **2. Twitter**

Twitter dio un giro de 180 grados: pasó de ser la plataforma que menos intervenía en los contenidos de los usuarios a la más activa. En marzo de 2020, avisó que ampliaría su definición de daño para incluir el contenido que vaya directamente en contra de las instrucciones de fuentes autorizadas en salud pública global y local.<sup>44</sup> Twitter prohibió entonces tuits que inviten a estas conductas o tengan los siguientes contenidos:

- Negar las recomendaciones de las autoridades con la intención de que las personas actúen en contra de estas.
- Alentar a romper el distanciamiento social.
- Recomendar tratamientos ineficaces, incluso si no son dañinos o si se comparten en broma.
- Recomendar tratamientos nocivos.
- Negar datos científicos sobre el contagio.
- Hacer afirmaciones que inciten a la acción, causen pánico, malestar o desorden a gran escala.
- Incluir información falsa o engañosa sobre el diagnóstico.
- Hacer afirmaciones de que grupos específicos o nacionalidades no son susceptibles o que son más susceptibles al virus.

La implementación de estas medidas no fue consistente. Mientras un par de tuits del entonces presidente de los Estados Unidos, Donald Trump, no merecieron

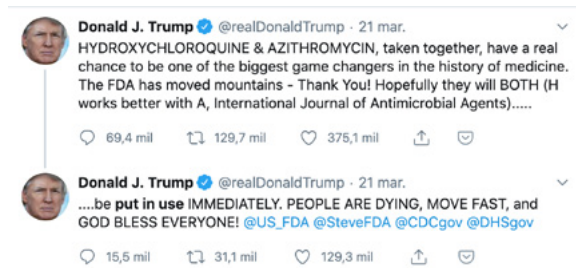
---

de realidad aumentada relacionados con el covid-19, a menos que hayan sido desarrollados en alianza con organizaciones de salud reconocidas. Jin, Kang-Xing, "Keeping People Safe and Informed about the Coronavirus", 18 de diciembre, 2020, disponible en: <https://about.fb.com/news/2020/10/coronavirus>, último acceso: 5 de marzo de 2021.

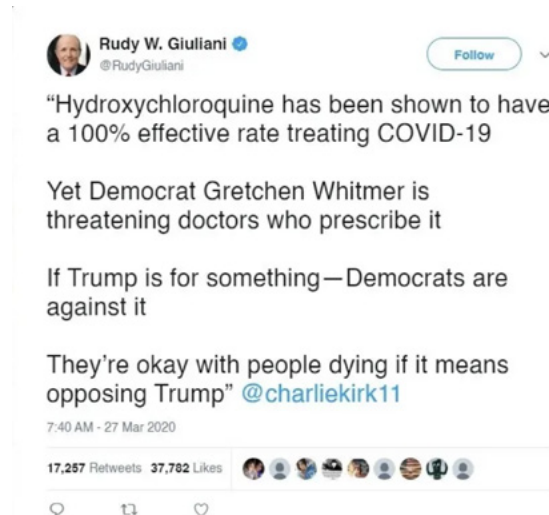
<sup>44</sup> Gadde, Vijaya y Derella, Matt, "Nueva información sobre nuestra estrategia continua sobre el covid-19", Twitter, 17 de marzo, 2020, disponible en: [https://blog.twitter.com/es\\_la/topics/company/2020/nueva-informacion-sobre-nuestra-estrategia-continua-sobre-el-covid-19.html](https://blog.twitter.com/es_la/topics/company/2020/nueva-informacion-sobre-nuestra-estrategia-continua-sobre-el-covid-19.html), último acceso: 5 de marzo de 2021.



ninguna respuesta de Twitter –decisión que, como veremos más adelante, estaría amparada en razones de interés público–, la plataforma sí actuó sobre un tuit del presidente de Brasil, Jair Bolsonaro, e incluso suspendió temporalmente la cuenta del abogado de Trump, Rudy Giuliani, por razones similares.<sup>45</sup>



Hilo publicado por el presidente Trump el 21 de marzo de 2020 donde promueve tratamientos médicos no comprobados contra el coronavirus.



Este tuit donde se promueve el uso de la hidroxiclороquina, publicado el 27 de marzo de 2020, fue eliminado por Twitter y le mereció a Giuliani la suspensión temporal de su cuenta.

En mayo de 2020, Twitter actualizó su enfoque sobre información engañosa relacionada con el covid-19 y aclaró que ejerce acciones sobre esos contenidos con base en tres categorías:<sup>46</sup>

- Información engañosa: declaraciones o afirmaciones que han sido confirmadas como falsas o engañosas por expertos en la materia, como lo son las autoridades de salud pública.
- Afirmaciones polémicas: declaraciones o afirmaciones en las que la preci-

<sup>45</sup> El presidente Bolsonaro publicó el 29 de marzo de 2020 dos videos que fueron eliminados por Twitter, en los que se le veía visitando lugares públicos en Brasilia, en los cuales alentaba a la gente a no confinarse y hablaba en favor del uso de la hidroxiclороquina para tratar el virus. La eliminación se dio días después del anuncio con los nuevos tipos de contenidos prohibidos. Darlington, Shasta, “Twitter elimina publicaciones sobre coronavirus del presidente de Brasil, Jair Bolsonaro”, CNN, 30 de marzo, 2020, disponible en: <https://cnnespanol.cnn.com/2020/03/30/twitter-elimina-publicaciones-sobre-coronavirus-del-presidente-de-brasil-jair-bolsonaro>, último acceso: 5 de marzo de 2021.

<sup>46</sup> Roth, Yoel y Pickles, Nick, “Actualizamos nuestro enfoque sobre información engañosa”, Twitter, 11 de mayo, 2020, disponible en: [https://blog.twitter.com/es\\_la/topics/product/2020/actualizamos-nuestro-enfoque-sobre-informacion-enganosa.html](https://blog.twitter.com/es_la/topics/product/2020/actualizamos-nuestro-enfoque-sobre-informacion-enganosa.html), último acceso: 5 de marzo de 2021.

sión, veracidad o credibilidad de la aseveración es realmente cuestionada o desconocida.

- **Afirmaciones no verificadas:** información que no está confirmada en el momento en que se comparte.<sup>47</sup>

Para cada tipo de contenido la propensión al daño puede ser moderada o severa. Según Twitter, la eliminación de publicaciones se hace en caso de información engañosa con una propensión al daño severa. Para los demás casos, se usan etiquetas y filtros:<sup>48</sup>

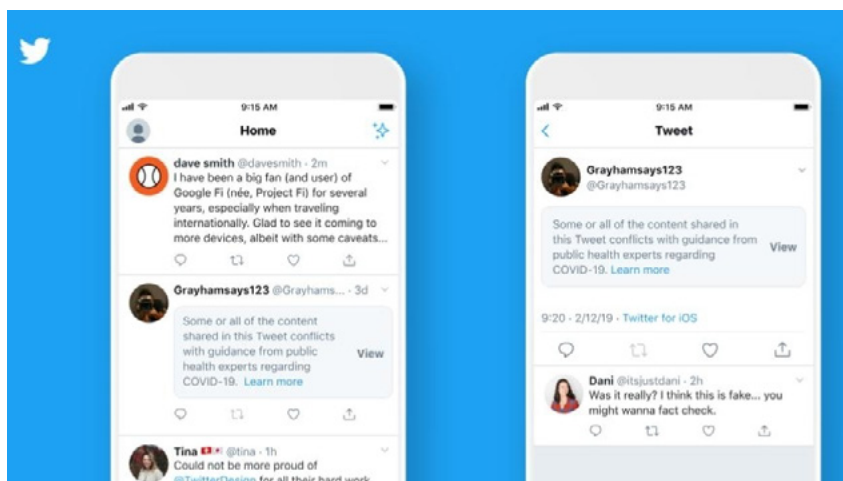
Información engañosa en Twitter		
Categorías \ propensión al daño	Moderada	Severa
Información engañosa	Filtro	Eliminación
Afirmaciones polémicas	Etiqueta	Filtro
Afirmaciones no verificadas	Ninguna acción	Ninguna acción en principio. Etiqueta conforme sea necesario.

Los filtros o avisos ocultan el tuit cuestionado para notificar a la persona de que el contenido difiere de la orientación de los expertos en salud pública. Siguiendo la explicación de la plataforma, estos se aplican en casos de información engañosa con propensión moderada al daño o afirmaciones polémicas con propensión severa al daño.

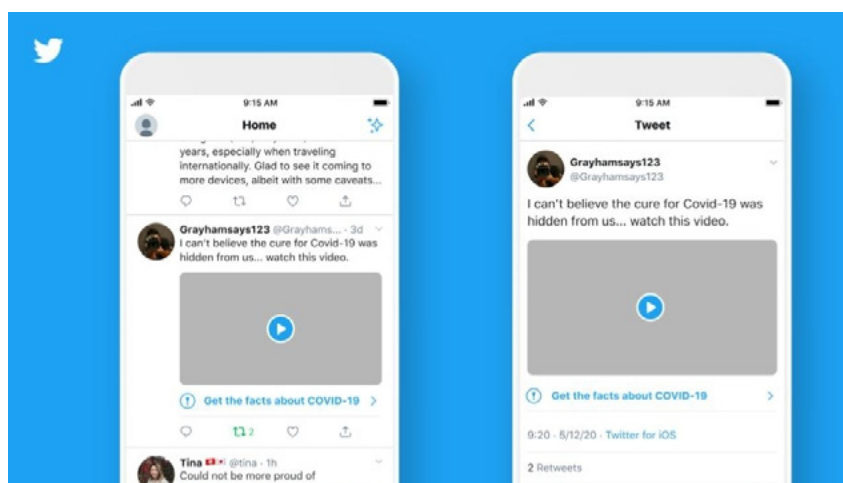
Las etiquetas, por su parte, aparecen como frases abajo del tuit, acompañadas de un signo de exclamación, que remiten a información confiable. Las etiquetas pueden utilizarse en casos de afirmaciones polémicas con propensión moderada al daño y en algunos casos de afirmaciones no verificadas con propensión severa al daño.

<sup>47</sup> *Ibíd.*

<sup>48</sup> La palabra que usa Twitter para los filtros es “aviso”. Yoel y Pickles, “Actualizamos nuestro enfoque sobre información engañosa”, *op. cit.*



Ejemplo de filtro o “aviso”.



Ejemplo de etiqueta.

Al igual que Facebook, Twitter anunció estas reglas en las entradas del blog de la compañía. Sin embargo, hasta la fecha no aparecen incorporadas en sus normas comunitarias, ni está claro si se implementarán de la misma manera en otros casos similares.

### 3. YouTube

YouTube, que ya tenía normas que prohibían la publicación de contenidos que recomienden el uso de sustancias peligrosas, creó una nueva “política sobre información médica errónea relacionada con el covid-19”.<sup>49</sup> Estas reglas no fueron

<sup>49</sup> Google, “Política sobre información médica errónea relacionada con el covid-19”, disponible en: <https://support.google.com/youtube/answer/9891785>, último acceso: 5 de marzo de 2021.

---

presentadas en forma de anuncios, sino que fueron directamente incorporadas en sus normas comunitarias. Allí se prohíbe el contenido sobre el covid-19 que implique un “riesgo grave de daño flagrante”. Según estas reglas, YouTube no permite el contenido que divulgue información médica errónea en la que se contradigan las indicaciones de la OMS o las autoridades de salud locales en relación con el tratamiento, la prevención, el diagnóstico o la transmisión del virus. Por ejemplo, se prohíbe afirmar que no ha habido fallecimientos a causa del covid o que hay una cura garantizada contra el virus; sostener que ciertas personas son inmunes al virus debido a su raza o nacionalidad; disuadir a los usuarios de consultar a un profesional médico si enferman o animarlos a recurrir a remedios caseros en lugar de buscar atención médica; afirmar que el origen del virus se encuentra en las redes 5G; declarar que el distanciamiento social no es una medida efectiva para reducir la propagación del virus; y manifestar que la vacuna contra el virus causará la muerte.

La infracción a la política da lugar al retiro del contenido. Además, YouTube aplica su sistema de faltas: con la primera infracción, el usuario recibe una advertencia. Luego de la primera vez YouTube añade una falta. Tres faltas dan lugar a la eliminación del canal. En esta política, YouTube también incluyó la excepción con fines educativos, documentales, científicos o artísticos. Según la política, es posible que se permita el contenido que en principio está prohibido si el contenido incluye contexto “en el que se concede el mismo peso o más a los puntos de vista equilibrados de las autoridades sanitarias locales o a los consensos científicos o médicos establecidos”. También se pueden hacer excepciones si la finalidad es rebatir desinformación, al proporcionar el contexto que lo justifique. El siguiente acápite explica más sobre las excepciones que aplican las plataformas a sus propias normas.

#### **IV. Las excepciones de “interés público” o similares**

Los estándares internacionales sobre derechos humanos defienden de manera especial las expresiones relacionadas con el interés público, lo cual implica que se les otorgan un mayor umbral de protección. La Corte Interamericana de Derechos Humanos vincula el interés público a asuntos relacionados con el funcionamiento de la democracia y el Estado, así como a la gestión pública y al ejercicio

de derechos.<sup>50</sup> De la misma forma, las redes sociales protegen de manera prevalente ciertos tipos de expresiones por considerarlas de interés público. Se trata, entonces, de la excepción que aplican las plataformas a contenidos que pueden ser violatorios de sus normas pero que deciden no remover por considerar que son del interés público de la ciudadanía.

Por lo general, el criterio para la utilización de la excepción se centra en el sujeto que produce el contenido. Si una persona dice que los niños son inmunes al coronavirus, esta publicación puede ser removida por violar las normas de Twitter. Pero la situación puede cambiar si quien lo expresa es Donald Trump, dado que los usuarios tienen un alto interés en informarse sobre las visiones del presidente de Estados Unidos.

Se trata de un criterio de difícil aplicación que en el caso de las redes sociales se vuelve aún más confuso. Como se vio, las plataformas no han concedido al presidente de Brasil la misma protección que al presidente Trump, a pesar de tratarse de expresiones de interés público. En últimas, la excepción le da un margen amplio de maniobra a las redes sociales para decidir caso a caso.

Si bien las tres compañías se han referido de una u otra forma a la defensa de expresiones de interés público en sus plataformas, ninguna parece tenerlo enteramente resuelto. La aplicación de las excepciones tampoco se incluye en los informes de transparencia que Facebook y YouTube difunden periódicamente.

## **1. Facebook**

“En un puñado de ocasiones cada año, dejamos en línea contenido que de otra manera violaría nuestras políticas si el valor del interés público sobrepasa el riesgo de daño. Frecuentemente, ver las expresiones de políticos es de interés público, y de la misma manera como los medios de comunicación reportan lo que dice un político, creemos que generalmente las personas deberían poder verlas por ellas mismas en nuestras plataformas”.<sup>51</sup> Así explicó Mark Zuckerberg la llama-

<sup>50</sup> “En cuanto al carácter de interés público, en su jurisprudencia la Corte ha reafirmado la protección a la libertad de expresión respecto de las opiniones o informaciones sobre asuntos en los cuales la sociedad tiene un legítimo interés de mantenerse informada, de conocer lo que incide sobre el funcionamiento del Estado, o afecta derechos o intereses generales o le acarrea consecuencias importantes”. Corte Interamericana de Derechos Humanos, “Caso Fontevecchia y D’Amico vs. Argentina”, sentencia del 29 de noviembre de 2011, § 61, disponible en: [https://corteidh.or.cr/docs/casos/articulos/seriec\\_238\\_esp.pdf](https://corteidh.or.cr/docs/casos/articulos/seriec_238_esp.pdf), último acceso: 5 de marzo de 2021.

<sup>51</sup> Traducción informal. “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

da *newsworthiness exemption* o excepción de interés noticioso.<sup>52</sup> Esta excepción no está establecida en las normas comunitarias de Facebook, sino que ha sido informada al público a través de anuncios en el blog de la compañía o en declaraciones públicas de sus representantes.<sup>53</sup>

En la aplicación de tal excepción, podrán mantenerse arriba las publicaciones que violen las normas de Facebook que tengan contenido noticioso, significativo o importante para el interés público. En septiembre de 2019, Facebook comunicó que, por regla general, las publicaciones de políticos están cobijadas por esta excepción –que, sin embargo, no aplica para contenido publicitado–.

Según Facebook, para evaluar si es necesario usar la excepción, se pondera el valor de interés público de la publicación versus el riesgo de daño, teniendo en cuenta estándares internacionales de derechos humanos. Si el daño es mayor que el interés del público en conocer el contenido, Facebook optaría por su retiro. De acuerdo con la red social, al examinar el interés público, se toman en cuenta factores como las circunstancias del país, como si hay una elección en marcha o si el país está en guerra; el contenido de lo dicho, incluyendo si se relaciona con la gobernanza o la política; y la estructura política del país, donde se mira si el país tiene una prensa libre.

Los límites de la excepción también se han hecho públicos a través de declaraciones de los representantes de Facebook. Nick Clegg, vicepresidente de asuntos globales y comunicaciones de Facebook, sostuvo que Facebook traza la raya cuando se trata de publicaciones que puedan llevar a violencia y a daños en el mundo real. Mark Zuckerberg también sostuvo en su página personal de Facebook que la excepción no es aplicable al contenido que incite a la violencia o busque suprimir votantes.<sup>54</sup> En esta ocasión, Zuckerberg también informó que la empresa empezaría a etiquetar los contenidos que deja en línea en aplicación de

---

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". Zuckerberg, Mark, Facebook, 26 de junio, 2020, disponible en: <https://www.facebook.com/zuck/posts/10112048980882521>, último acceso: 5 de marzo de 2021.

<sup>52</sup> Kaplan, Joel y Osofsky, Justin, "Input from Community and Partners on our Community Standards", Facebook, 21 de octubre, 2016, disponible en: <https://about.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards>, último acceso: 5 de marzo de 2021, y Clegg, Nick, "Facebook, Elections and Political Speech", 24 de septiembre, 2019, disponible en: <https://about.fb.com/news/2019/09/elections-and-political-speech>, último acceso: 5 de marzo de 2021.

<sup>53</sup> En un discurso dictado en la Universidad de Georgetown en 2019, Mark Zuckerberg declaró que no creía que en una democracia esté bien que una empresa privada censure a los políticos o las noticias. Facebook, "Mark Zuckerberg Stands for Voice and Free Expression", 17 de octubre, 2019, disponible en: <https://about.fb.com/news/2019/10/mark-zuckerberg-stands-for-voice-and-free-expression>, último acceso: 5 de marzo de 2021.

<sup>54</sup> Zuckerberg, Facebook, *op. cit.*



la excepción y que permitiría que las personas los compartan para condenarlos.<sup>55</sup>

Sin embargo, la utilización en la práctica deja dudas. En septiembre de 2020, Facebook borró por violación a su política en contra de la incitación a la violencia una imagen de una candidata (posteriormente elegida) al Congreso de los Estados Unidos por el estado de Georgia en la que posaba con un rifle junto a tres fotografías de políticos demócratas.<sup>56</sup> En ese mismo mes, Facebook borró una publicación de un congresista de Luisiana en el que prometía el uso de fuerza letal en contra de manifestantes.<sup>57</sup> En contraste, la plataforma mantuvo en línea un mensaje del presidente Trump en la que sostuvo, casi de manera amenazante, que “cuando comienza el saqueo, comienza el tiroteo” también en el contexto de una manifestación pública.<sup>58</sup>

El covid-19 también ha hecho evidente las inconsistencias en el empleo de la política: un video de Jair Bolsonaro en el que hablaba a favor del consumo de la hidroxiclороquina en el tratamiento del virus fue eliminado de Facebook en marzo de 2020,<sup>59</sup> pero una publicación de Trump en el mismo sentido se mantuvo en línea.<sup>60</sup> No obstante, otro mensaje en el que Trump compara al covid-19 con la influenza sí fue eliminado de Facebook.<sup>61</sup>

## 2. Twitter

La excepción de interés público está establecida de manera más clara en las normas comunitarias de Twitter.<sup>62</sup> Según estas, un contenido es de interés público

---

<sup>55</sup> *Ibid.*

<sup>56</sup> Associated Press, “Georgia Candidate’s Post Removed; Facebook Says It Violates Policy against Inciting Violence”, Wabe, 4 de septiembre, 2020, disponible en: <https://www.wabe.org/georgia-candidates-post-removed-facebook-says-it-violates-policy-against-inciting-violence>, último acceso: 5 de marzo de 2021.

<sup>57</sup> Associated Press, “Facebook Removes Congressman’s Post over ‘Incitement’”, ABC News, 2 de septiembre, 2020, disponible en: <https://abcnews.go.com/Politics/wireStory/facebook-removes-congressmans-post-incitement-72778776>, último acceso: 5 de marzo de 2021.

<sup>58</sup> “When the looting starts, the shooting starts”. Trump, Donald, Facebook, 28 de mayo, 2020, disponible en: <https://www.facebook.com/DonaldTrump/posts/10164767134275725>, último acceso: 5 de marzo de 2021.

<sup>59</sup> Constine, Josh, “Facebook Deletes Brazil President’s Coronavirus Misinfo Post”, Tech Crunch, 30 de marzo, 2020, disponible en: <https://techcrunch.com/2020/03/30/facebook-removes-bolsonaro-video>, último acceso: 5 de marzo de 2021.

<sup>60</sup> Trump, Donald, Facebook, 21 de marzo, 2020, disponible en: <https://www.facebook.com/DonaldTrump/posts/10164254051615725>, último acceso: 5 de marzo de 2021.

<sup>61</sup> Ingram, David, “Facebook Removes Trump Post that Compared Covid-19 to Flu”, NBC News, 6 de octubre, 2020, disponible en: <https://www.nbcnews.com/tech/tech-news/facebook-removes-trump-post-compared-covid-19-flu-n1242277>, último acceso: 5 de marzo de 2021.

<sup>62</sup> Twitter, “Acerca de las excepciones de interés público en Twitter”, disponible en: <https://help.twitter.com/es/rules-and-po>

“si constituye un aporte directo para la comprensión o el debate de un asunto que le preocupa a todo el público”. Actualmente, las excepciones solo se aplican a tuits de funcionarios electos y del Gobierno, y candidatos a cargos políticos.<sup>63</sup> En caso de mantener en línea un tuit amparado en esta excepción, Twitter agrega un aviso o filtro que proporciona contexto sobre el incumplimiento de las reglas. Esto también hace que el tweet no sea recomendado por el algoritmo de Twitter y limita la capacidad de los usuarios de interactuar con lo publicado.

La plataforma pondera el interés público del contenido frente al posible riesgo y a la gravedad del daño. A diferencia de Facebook y YouTube, Twitter detalla en sus normas los factores que considera para tomar sus decisiones:<sup>64</sup>

- Hay violaciones en las que es más probable que se aplique la excepción (por ejemplo, en los casos de incitación al odio o acoso) y violaciones en las que es menos probable que se aplique y, en consecuencia, se elimina el contenido (como en casos de terrorismo, violencia o integridad electoral).
- Hay criterios que hacen más probable que se aplique la excepción, como cuando el tuit está dirigido a funcionarios del Gobierno o cuando proporciona un contexto importante para eventos geopolíticos en curso. Así mismo, es menos probable que se aplique la excepción cuando, por ejemplo, el tuit incluye un llamado a la acción.
- En ningún caso se hacen excepciones para contenido multimedia relacionado con la explotación sexual infantil, la desnudez no consensuada y la agresión sexual violenta a las víctimas.

En octubre de 2019, Twitter se refirió a la aplicación de la excepción para los líderes mundiales, al explicar que es importante asegurar el derecho de la gente a saber de sus líderes y exigir rendición de cuentas.<sup>65</sup> Twitter aclaró que estos no se encuentran por encima de sus reglas enteramente y que, en caso de utilizar la

---

licies/public-interest, último acceso: 5 de marzo de 2021. También son relevantes dos publicaciones en el blog de Twitter: Twitter Safety, “Defining Public Interest on Twitter”, 27 de junio, 2019, disponible en: [https://blog.twitter.com/en\\_us/topics/company/2019/publicinterest.html](https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html), último acceso: 5 de marzo de 2021, y Twitter Inc., “World Leaders on Twitter: Principles & Approach”, 15 de octubre, 2019, disponible en: [https://blog.twitter.com/en\\_us/topics/company/2019/worldleaders2019.html](https://blog.twitter.com/en_us/topics/company/2019/worldleaders2019.html), último acceso: 5 de marzo de 2021.

<sup>63</sup> Para aplicar la excepción, el tweet tiene que ser publicado desde una cuenta verificada que tenga por lo menos cien mil seguidores.

<sup>64</sup> Twitter publica listas más detalladas, aquí se presentan algunos ejemplos.

<sup>65</sup> Twitter Inc., “World Leaders on Twitter: Principles & Approach”, *op. cit.*

excepción, puede ponerse un filtro. En noviembre de 2020, Jack Dorsey sostuvo, en una audiencia ante el Congreso de los Estados Unidos, que Donald Trump no tendría estas protecciones cuando finalice su mandato.<sup>66</sup> “Si una cuenta de repente ya no es de un líder mundial, esa política en particular se va”, dijo Dorsey. Sin embargo, es cuestionable afirmar que se pierde el estatus de líder mundial al salir de la Casa Blanca.



Twitter aplicó la excepción a este tuit de Donald Trump por violación a sus normas de sobre desinformación y covid. Antes de poder ver el contenido, aparece un intersticial con la advertencia.<sup>67</sup>

### 3. YouTube

Las normas de la comunidad de YouTube no establecen propiamente una excepción de interés público. La única mención de una excepción de ese estilo se encuentra en una declaración de Susan Wojcicki, directora ejecutiva de YouTube, quien sostuvo que la plataforma podría conservar contenido publicado por políticos que fuera violatorio de sus normas: “cuando tienes a un político que está produciendo información que es realmente importante que vean sus electores, ese es un contenido que dejaríamos arriba porque pensamos que es importante que otra gente lo vea”.<sup>68</sup>

Después de estas palabras, un vocero de YouTube declaró que los políticos no son tratados de manera diferente por la plataforma, sino que se les conceden excepciones a ciertas clases de discurso con contenido educativo, documental, cientí-

<sup>66</sup> Hamilton, Isobel Asher, “Trump Will Lose his ‘World Leader’ Twitter Privileges on January 20, Jack Dorsey Confirms”, Insider, 18 de noviembre, 2020, disponible en: <https://www.businessinsider.com/donald-trump-twitter-account-lose-world-leader-protections-exemption-20-2020-11>, último acceso: 5 de marzo de 2021.

<sup>67</sup> Twitter, 11 de octubre, 2020, disponible en: <https://twitter.com/realDonaldTrump/status/1315316071243476997>, último acceso: 5 de marzo de 2021.

<sup>68</sup> Traducción informal. “When you have a political officer that is making information that is really important for their constituents to see, or for other global leaders to see, that is content that we would leave up because we think it’s important for other people to see”. Overly, Steven, “YouTube CEO: Politicians Can Break our Content Rules”, Politico, 25 de septiembre, 2019, disponible en: <https://www.politico.com/story/2019/09/25/youtube-ceo-politicians-break-content-rules-1510919>, último acceso: 5 de marzo de 2021.

fico o artístico –que YouTube reúne bajo la sigla EDSA–.<sup>69</sup> Entre otros, el criterio EDSA aparece en políticas sobre contenido perjudicial o peligroso, contenido violento o gráfico, incitación al odio o a la violencia y, como se mencionó antes, en la política sobre información médica errónea relacionada con el covid-19. No obstante, más que una excepción que permita mantener en línea contenido infractor, el criterio EDSA enfatiza la necesidad de proveer contexto para entender la intención del video.<sup>70</sup>

## V. Nuevas reglas para la publicidad

Las tres plataformas han tomado varias medidas en relación con la publicidad paga en sus plataformas. Vale la pena señalar que, en lo que respecta al contenido publicitado, la lógica es distinta: las plataformas no permiten por regla general cualquier publicación a la espera de que luego se identifiquen posibles violaciones a sus normas, como ocurre con el contenido orgánico. Por el contrario, los contenidos que se quieren publicitar son previamente presentados por los contratantes y aprobados por las plataformas. Es por esto que en estos casos no se habla de eliminación de contenidos sino de contenidos no autorizados.

### 1. Contenido no autorizado: elementos de bioseguridad

Las tres plataformas han tomado varias medidas respecto del contenido publicitado que pueda afectar la disponibilidad de elementos de bioseguridad, incluso si los anuncios no son desinformativos.<sup>71</sup> En febrero, antes de la declaratoria de la pandemia, Facebook prohibió los avisos que buscaran crear pánico o que denotaran urgencia respecto a suministros y productos vinculados con el covid-19. En términos de contenido falso, la red social vedó los anuncios que garanticen la cura o la prevención del virus. En marzo, Facebook prohibió las ventas de tapabocas, geles, toallas desinfectantes y kits de prueba para el covid-19.<sup>72</sup> Esta última prohibición se ha ido relajando: desde agosto de 2020, Face-

---

<sup>69</sup> *Ibid.*

<sup>70</sup> Google, “La importancia del contexto”, disponible en: <https://support.google.com/youtube/answer/6345162>, último acceso: 5 de marzo de 2021.

<sup>71</sup> Este documento muestra las medidas tomadas al 30 de noviembre de 2020.

<sup>72</sup> Jin, “Keeping People Safe and Informed about the Coronavirus”, *op. cit.*, “Banning Ads for Hand Sanitizer, Disinfecting Wipes and Covid-19 Testing Kits”, actualización del 19 de marzo, 2020, 2:18 PM.

book ya permite la promoción de tapabocas no médicos (bajo el cumplimiento de ciertos requisitos)<sup>73</sup> y de geles y toallas desinfectantes.<sup>74</sup>

De manera similar, en abril de 2020 Twitter prohibió el contenido sensacionalista o que incite al pánico y anuncios con los precios inflados. Así mismo, prohibió la venta de tapabocas y geles y no permitió menciones de vacunas, tratamientos o kits de prueba, excepto que sea información publicada por medios que la plataforma exceptúa bajo su política de anuncios de carácter político.<sup>75</sup> Posteriormente, en agosto de 2020, se limitó la prohibición a los tapabocas médicos.

Finalmente, al igual que Facebook, Google restringe las ventas de tapabocas requeridos por trabajadores de la salud, para evitar la escasez de los suministros, y permite únicamente las de tapabocas elaborados exclusivamente con tela.<sup>76</sup> La empresa también comunicó que está tomando medidas para evitar los aumentos artificiales de precios.

	Menciones <sup>77</sup>	Ventas	Otros
FB		Tapabocas médicos	Pánico/urgencia
		Geles	Cura/prevención del virus
		Toallas desinfectantes	
		Kits de prueba	
TW	Vacunas	Tapabocas	Pánico/sensacionalista
	Tratamientos	Geles	Precios artificiales
	Kits de prueba		
YT		Tapabocas médicos	Precios artificiales

## 2. Contenido no autorizado: términos relacionados con covid-19

Además de las restricciones relacionadas con elementos de bioseguridad, las plataformas tomaron otras medidas para el contenido publicitado. Tanto Twitter

<sup>73</sup> Leathern, Rob, "Allowing the Promotion of Non-Medical Masks on Facebook", Facebook, 10 de junio, 2020, disponible en: <https://www.facebook.com/business/news/allowing-the-promotion-of-non-medical-masks-on-facebook>, último acceso: 5 de marzo de 2021.

<sup>74</sup> Jin, "Keeping People Safe and Informed about the Coronavirus", *op. cit.*, actualización del 19 de agosto de 2020, 10:05 AM.

<sup>75</sup> Twitter, "Contenido de carácter político", disponible en: <https://business.twitter.com/es/help/ads-policies/ads-content-policies/political-content.html>, último acceso: 5 de marzo de 2021.

<sup>76</sup> Google, "Actualizaciones de la política de Google Ads sobre la enfermedad del coronavirus (covid-19)", disponible en: <https://support.google.com/google-ads/answer/9811449?hl=es-419>, último acceso: 5 de marzo de 2021.

<sup>77</sup> Actualizado al 30 de noviembre de 2020.

como Google iniciaron la pandemia con una prohibición general para cualquier publicidad que usara palabras relacionadas con el covid-19. Siendo así, no se permitieron publicaciones sobre asuntos de interés público o que podrían ser útiles para los usuarios –como canales de donación o seguros de salud–. Según Google, esto se decidió al aplicar su política de contenido inapropiado, existente antes de la pandemia.<sup>78</sup> Según esta política, no se permiten anuncios que se lucren potencialmente de un “evento delicado”, como una catástrofe natural o un conflicto. Con la pandemia, Google actualizó la política para incluir las “emergencias de salud pública” como un tipo de evento delicado.<sup>79</sup>

La decisión de Google fue fuertemente cuestionada por miembros del Partido Demócrata de los Estados Unidos.<sup>80</sup> Dado que la prohibición tenía excepciones para entidades del Gobierno, la administración de Donald Trump estaba habilitada para publicar anuncios en relación con su respuesta a la pandemia, mientras que los demócratas no tenían la oportunidad de presentar publicidad crítica de las acciones del Gobierno frente a la crisis.

En abril, estas compañías introdujeron algunas modificaciones.<sup>81</sup> Además de prohibir las ventas de elementos de bioseguridad, Twitter prohibió las referencias publicitarias de mal gusto sobre el covid-19.<sup>82</sup> Google, por su parte, informó que permitiría anuncios de “proveedores de atención médica, organizaciones gubernamentales, no gubernamentales e intergubernamentales, anunciantes que publiquen mensajes electorales verificados y cuentas administradas del sector privado con un historial de cumplimiento de políticas que quieran compartir información relevante con el público”.<sup>83</sup>

En diciembre de 2020, al prever el inicio de la vacunación contra el covid-19 en todo el

---

<sup>78</sup> Pichai, Sundar, “Coronavirus: cómo estamos ayudando”, Google, 6 de marzo, 2020, disponible en: <https://blog.google/inside-google/company-announcements/coronavirus-covid19-como-estamos-ayudando>, último acceso: 5 de marzo de 2021, y Google, “Contenido inapropiado”, disponible en: <https://support.google.com/adspolicy/answer/6015406?hl=es-419>, último acceso: 5 de marzo de 2021.

<sup>79</sup> Google, “Actualizaciones de la política de Google Ads sobre la enfermedad del coronavirus (covid-19)”, *op. cit.*

<sup>80</sup> Birnbaum, Emily, “Democrats Say Google’s Covid-19 Ad Ban Is a Gift to Donald Trump”, Protocol, 31 de marzo, 2020, disponible en: <https://www.protocol.com/google-coronavirus-ad-ban-democrats>, último acceso: 5 de marzo de 2021.

<sup>81</sup> Fischer, Sara, “Twitter Lifts Coronavirus Ad Ban”, Axios, 3 de abril, 2020, disponible en: <https://www.axios.com/twitter-coronavirus-ad-ban-eb9c8946-d90b-4bb2-be29-9fba531e44d8.html>, último acceso: 5 de marzo de 2021, y Birnbaum, Emily, “Google Revises Covid-19 Ad Ban after Backlash”, Protocol, 2 de abril, 2020, disponible en: <https://www.protocol.com/google-coronavirus-ad-ban-reverse>, último acceso: 5 de marzo de 2021.

<sup>82</sup> Ad Exchanger, “Google, Twitter Lift Coronavirus Ad Ban; Facebook Offers Grants to SMBs”, 6 de abril, 2020, disponible en: <https://www.adexchanger.com/ad-exchange-news/monday-06042020>, último acceso: 5 de marzo de 2021.

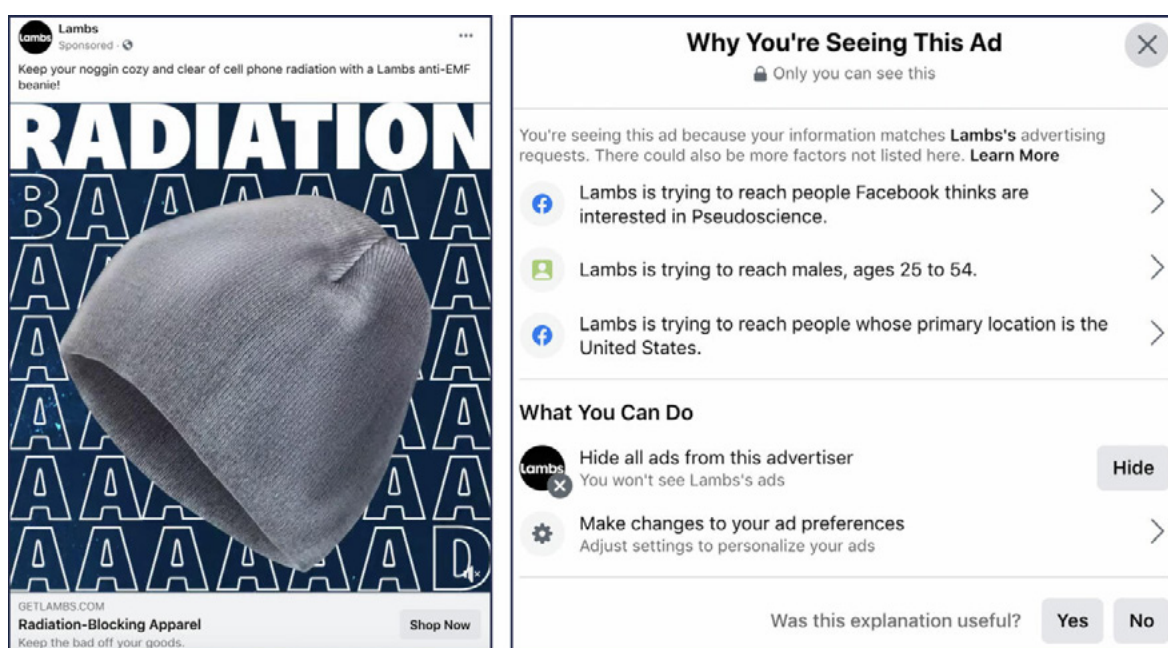
<sup>83</sup> Google, “Actualizaciones de la política de Google Ads sobre la enfermedad del coronavirus (covid-19)”, *op. cit.*



mundo, Facebook comunicó que prohibiría el contenido publicitado donde se vendieran kits de vacunación o se promocionara el acceso acelerado a la vacuna. Facebook también prohíbe los anuncios que sostengan que la vacuna es una cura para el virus.<sup>84</sup>

### 3. Eliminación de la categoría "pseudociencia"

En abril de 2020, el medio de comunicación The Markup mostró cómo era posible contratar publicidad dirigida a personas que, según Facebook, estuvieran interesadas en "pseudociencia".<sup>85</sup> Esta categoría contenía más de 78 millones de personas. Luego de la publicación de The Markup, Facebook eliminó silenciosamente esta categoría como posibilidad para la compra de anuncios dirigidos, lo que en últimas impactaba en una posible publicidad alrededor del covid-19.<sup>86</sup>



The Markup encontró un anuncio para un gorro que dice proteger la cabeza de la radiación celular. Las preocupaciones por la radiación electromagnética emitida por la infraestructura de la telefonía móvil 5G hacen parte de las teorías de conspiración relacionadas con el coronavirus. Al dar clic en "¿Por qué estás viendo este anuncio?", Facebook mostró que la empresa anunciante quería llegar a personas que puedan estar interesadas en pseudociencia.<sup>87</sup>

<sup>84</sup> Jin, "Keeping People Safe and Informed about the Coronavirus", *op. cit.*

<sup>85</sup> Sankin, Aaron, "Want to Find a Misinformed Public? Facebook's already Done It", The Markup, 23 de abril, 2020, disponible en: <https://themarkup.org/coronavirus/2020/04/23/want-to-find-a-misinformed-public-facebooks-already-done-it>, último acceso: 5 de marzo de 2021.

<sup>86</sup> Culliford, Elizabeth, "Facebook Gets Rid of 'Pseudoscience' Ad-targeting Category", Reuters, 23 de abril, 2020, disponible en: <https://uk.reuters.com/article/health-coronavirus-facebook-ads/facebook-gets-rid-of-pseudoscience-ad-targeting-category-idUKL2N2CB1D6?feedType=RSS&feedName=technology-media-telco-SP>, último acceso: 5 de marzo de 2021.

<sup>87</sup> Sankin, "Want to Find a Misinformed Public? Facebook's already Done It", *op. cit.*

## 4. Desmonetización

Google también tomó medidas que afectaron a otro grupo de actores de su ecosistema de anuncios: los creadores del programa de socios de YouTube. Como se mencionó anteriormente, YouTube da una porción de las ganancias de la publicidad a las cuentas que forman parte de este programa.<sup>88</sup> Sin embargo, al inicio de la pandemia, la compañía avisó que no mostraría publicidad (lo que se conoce como “desmonetizar”) en los videos enfocados en el covid-19.<sup>89</sup> Esto también lo hizo al extender su política de “eventos delicados”. Esta medida también fue flexibilizada. En marzo, YouTube dio aviso en su blog que permitiría anuncios para videos que discutan el covid-19 en un número limitado de canales.<sup>90</sup>

## VI. Conclusiones

Pasará mucho tiempo antes de que comprendamos el impacto que ha tenido la pandemia del coronavirus en la humanidad y, en particular, los efectos de este evento global en el ejercicio de la libertad de expresión en internet. No obstante, una consecuencia que podemos aventurar en este último campo es el profundo cambio en la moderación de contenidos de las redes sociales.

Sería equivocado afirmar, en cualquier caso, que este nuevo escenario se dio únicamente por cuenta del covid-19. Las plataformas venían confrontando presiones políticas, regulatorias y sociales de todo tipo para intervenir de manera más decidida en los contenidos problemáticos de los usuarios. Para no ir lejos, 2020 también fue el año de la elección presidencial en Estados Unidos: la primera bandera roja para las redes sociales era que no se repitiera la foto de 2016.

Esta pandemia emergió en una sociedad globalizada que habita un entorno digital complejo, donde la interconexión ofrece a la par soluciones y problemas: por las mismas redes donde se empezaron a difundir las primeras alertas y medidas de salud pública circulaban rumores falsos y curas milagrosas. En ese contex-

---

<sup>88</sup> Se requieren por lo menos mil suscriptores y cuatro mil horas de reproducción para formar parte. Google, “Descripción general y elegibilidad del programa de socios de YouTube”, *op. cit.*

<sup>89</sup> Barca, Kamila, “Youtube ha desmonetizado los videos de influencers que hablan sobre el coronavirus”, Business Insider, 18 de febrero, 2020, disponible en: <https://www.businessinsider.es/youtube-ha-desmonetizado-todos-videos-coronavirus-583723>, último acceso: 5 de marzo de 2021.

<sup>90</sup> Wojcicki, Susan, “Coronavirus: An Update on Creator Support and Resources”, Blog Oficial de YouTube, 11 de marzo, 2020, disponible en: <https://blog.youtube/inside-youtube/coronavirus-update-on-creator-support>, último acceso: 5 de marzo de 2021.

---

to, servicios como Facebook, YouTube o Twitter –que ya se enfrentaban a varios interrogantes por su responsabilidad en la configuración del debate público– se enfocaron en salvar su propio pellejo mientras preservaban su modelo de negocio. El objetivo principal era evitar a toda costa que sus plataformas se convirtieran en un vector de desinformación y desconfianza tan contagioso como el coronavirus.

En los meses sucesivos habrá un debate sobre las soluciones por las que optaron estas plataformas. Por ahora, teniendo presente la excepcionalidad del problema, pero sin olvidar el camino que se venía recorriendo, planteamos algunas reflexiones y conclusiones:

- **La pandemia aterrizó en las redes sociales en medio de muchas preguntas sobre la moderación de contenidos.** Si bien este hecho acrecentó el problema, ya había un estado general de dispersión, ambigüedad e inconsistencia frente a cuáles son las normas comunitarias a las que están sujetos los usuarios, dónde se encuentran y de qué manera se implementan. Muchas reglas se anunciaron en las entradas de blog o en las noticias sin que llegaran a incorporarse formalmente.
- **¿Cuál es la naturaleza de las normas comunitarias?** El objetivo de este trabajo no era hacer un análisis hermenéutico de estos textos. Sin embargo, hay una duda fundamental sobre cuál es la naturaleza de estas reglas. En algunos casos las plataformas parecen postularlas de manera taxativa y en otros como la enunciación o la formulación de un principio sujeto a interpretación y desarrollo. De una u otra forma, la primera y la última palabra la tienen las plataformas.
- **La tormenta perfecta del covid-19 y la campaña presidencial norteamericana.** El objetivo principal de las redes sociales de cara a las elecciones de 2020 en los Estados Unidos era evitar la instrumentalización de las plataformas que se dio en 2016: el acceso a las bases de datos, la segmentación de la publicidad y las operaciones conjuntas influyeron en que hubiera un debate manipulado, altamente polarizado y, en esencia, desinformado. Con la llegada del coronavirus, la emergencia sanitaria y la electoral se encontraron, lo que puso de presente las limitaciones que enfrentan estas empresas para controlar a los actores políticos que capitalizan estrategias de desinformación –empezando por Donald Trump–.
- **La dificultad para designar autoridades confiables u oficiales.** La pandemia ha sido también una crisis de confianza. Con buena intención, las plataformas

---

intentaron darles mayor relevancia a las fuentes autorizadas de salud pública. No obstante, no era claro el proceso mediante el cual se daba ese voto de confianza ni tampoco había una forma de resolver las contradicciones evidentes entre las mismas autoridades de un país –como fue el caso de Brasil–.

- **Interés público sí, pero no.** La dificultad de trazar criterios claros para definir la excepción de interés público fue más patente en esta coyuntura. Durante esta crisis muchos líderes políticos difundieron mensajes problemáticos –como curas milagrosas– que, empero, tenían una connotación de interés público. Al final, lo que pareció prevalecer en la aplicación de la excepción fue evitar una crisis reputacional para las plataformas.
- **Decisiones de moderación en medio del debate político.** La manera como se implementó la excepción de interés público frente al presidente Donald Trump pone de presente el momento político en el que se debía tomar esa decisión y –en últimas, la visión de las plataformas– que se traducían en formas distintas de moderar el contenido. Twitter, que empezó a trazarle una línea roja a Trump, fue la primera en etiquetar sus tuits y, aunque aplicaba la excepción de interés público, con la medida redujo sustancialmente la posibilidad de interactuar con el contenido infractor. Facebook, por su parte, se mostró más reticente a meterse con el presidente, pero al utilizar la excepción y etiquetar el contenido aún permite las interacciones.
- **Entre la forma y el fondo.** Cada plataforma ha hecho énfasis distintos en la manera en que presenta las normas comunitarias en esta emergencia. Mientras Facebook ofrece más detalles en la parte sustancial, Twitter enfatiza los procesos. De las tres, YouTube suele encontrar un balance entre ambos puntos.
- **Mayor claridad en las normas para anunciantes.** Pareciera que las plataformas se han preocupado más en delinear claramente las reglas que deben seguir los anunciantes durante la pandemia que hacer lo propio con las normas comunitarias de los usuarios.

La intervención decisiva e imperfecta de las plataformas durante la pandemia abre una nueva etapa en la discusión sobre moderación de contenidos. Es posible que a medida que regrese la normalidad algunas reglas se revertan o maten –como, en efecto, ya ha pasado–, pero es difícil pensar que estos intermediarios puedan regresar a la visión anterior (enfocada, por ejemplo, en la actividad inauténtica

antes que en los contenidos) y a la narrativa de no querer ser árbitros de la verdad. El 2020 para las redes sociales coincidió también con la entrada en funcionamiento de la junta de supervisión de Facebook y con varias iniciativas de la sociedad civil que intentan influir en el debate sobre la responsabilidad de los intermediarios en la configuración del debate público digital: entre otros, el boicot contra anunciantes “Stop Hate for Profit”;<sup>91</sup> la campaña para reducir el odio en línea “Change the Terms”;<sup>92</sup> la actualización de los Principios de Santa Clara<sup>93</sup> sobre transparencia y rendición de cuentas en la moderación de contenidos; y, desde América Latina, la propuesta que lidera Observacom sobre estándares para la regulación de las plataformas.<sup>94</sup>

Estas iniciativas han querido centrar la conversación en preguntas clave sobre las normas comunitarias: ¿cuáles son los estándares que deben guiarlas? ¿Cómo hacerlas compatibles con los derechos humanos? ¿Cómo lograr que haya rendición de cuentas y alguna forma de “debido proceso”? ¿Cómo garantizar la vigencia de la libertad de expresión? Si bien se trata de cuestiones fundamentales en la agenda de los derechos humanos en la era digital, resolver estas preguntas no terminaría de iluminar el dilema práctico de la moderación de contenidos –cuestiones como la escala, la consistencia o la oportunidad– ni las tensiones económicas, sociales y políticas que se desarrollan en las redes sociales y que la pandemia exacerbó. Tal vez el eslogan desgastado del año 2020 excepcional aplique también para esta cuestión: es hora de reinventar esta conversación.

## Anexos

Las siguientes tablas resumen las medidas tomadas por Facebook, Twitter y YouTube entre febrero y noviembre de 2020, de acuerdo con lo explicado en acápites anteriores. Se señalan las diferencias entre el contenido orgánico y el contenido publicitado. Adicionalmente, para cada acción se señala si, hasta donde fue posible constatar, la medida se tomó en aplicación de una norma comunitaria previa-

---

<sup>91</sup> Stop Hate for Profit, disponible en: <https://www.stophateforprofit.org>, último acceso: 5 de marzo de 2021.

<sup>92</sup> Change the Terms, disponible en: <https://www.changetheterms.org>, último acceso: 5 de marzo de 2021.

<sup>93</sup> Santa Clara Principles, disponible en: <https://santaclaraprinciples.org>, último acceso: 5 de marzo de 2021.

<sup>94</sup> Libertad de Expresión y Plataformas de Internet, “Estándares para una regulación democrática de las grandes plataformas que garantice la libertad de expresión en línea y una Internet libre y abierta”, julio, 2020, disponible en: <https://www.observacom.org/wp-content/uploads/2020/09/Estandares-para-una-regulacion-democratica-de-las-grandes-plataformas.pdf>, último acceso: 5 de marzo de 2021.

mente establecida, si fue comunicada a través del blog de la compañía o si se trató de un cambio incorporado directamente en el texto de las normas comunitarias.

En términos generales, las medidas tomadas consisten en:

- Identificar contenidos tildados de falsos por *fact-checkers* o autoridades sanitarias, para eliminarlos, etiquetarlos o reducir su circulación.
- Enviar notificaciones a personas que planeen compartir o que hayan estado en contacto con contenidos señalados como falsos.
- Restringir los anuncios pagos y las publicaciones orgánicas con las que se pretenda vender ciertos elementos de bioseguridad (como tapabocas o geles antibacteriales), que busquen crear pánico o que aseguren curar o prevenir el virus.

Facebook				
Tipo de contenido	Medida	¿Aplicación de norma ya vigente?	¿Medida anunciada en el blog?	¿Cambio incorporado como nueva norma comunitaria?
	Eliminación de contenidos contrarios a autoridades de salud	Discutible	Sí	No
	Identificación de duplicados	Poco claro	Sí	No
	Prohibición de ventas orgánicas de ciertos elementos de bioseguridad	No	Sí	No
	Etiquetas y limitación de difusión de contenidos disputados por fact-checkers	Sí	Sí	N/A
Orgánico	Notificaciones de exposición a contenidos de desinformación	No	Sí	No
	[Instagram] Bloqueo o restricción en el uso de hashtags	No	Sí	No
	[Instagram] Eliminación de recomendados relacionados con el covid-19	No	Sí	No
	[Instagram] Deshabilitación de la opción de buscar efectos de realidad aumentada relacionados con el covid-19	No	Sí	No
Publicidad	Restricciones en la venta de ciertos elementos de bioseguridad y de anuncios que busquen crear pánico o que garanticen la cura o la prevención del virus	No	Sí	Sí
	Eliminación de la categoría "pseudociencia"	N/A	No	No
	Restricciones en la venta de kits de vacunas o en la promoción del acceso acelerado a ellas	No	Sí	Sí



<b>Twitter</b>				
<b>Tipo de contenido</b>	<b>Medida</b>	<b>¿Aplicación de norma ya vigente?</b>	<b>¿Medida anunciada en el blog?</b>	<b>¿Cambio incorporado como nueva norma comunitaria?</b>
	Eliminación de tweets contrarios a las autoridades de salud	No	Sí	No
	Eliminación de información engañosa con una propensión al daño severa	No	Sí	No
Orgánico	Filtros para información engañosa con propensión moderada al daño o a las afirmaciones polémicas con propensión severa al daño	No	Sí	No
	Etiquetas para afirmaciones polémicas con propensión moderada al daño y en algunos casos de afirmaciones no verificadas con propensión severa al daño	No	Sí	No
Publicidad	Restricciones en la venta de ciertos elementos de bioseguridad	No	Sí	Sí
	Prohibición general para palabras relacionadas con el covid-19	No	Sí	Sí

<b>YouTube</b>				
<b>Tipo de contenido</b>	<b>Medida</b>	<b>¿Aplicación de norma ya vigente?</b>	<b>¿Medida anunciada en el blog?</b>	<b>¿Cambio incorporado como nueva norma comunitaria?</b>
	Prohibición de recomendaciones de sustancias peligrosas o tratamientos dañinos	Sí	N/A	N/A
Orgánico	Prohibición de contenido sobre covid-19 que implique un riesgo de daño flagrante	No	Sí	Sí
	Prohibición de contenido contrario a autoridades de salud	No	Sí	Sí
Publicidad	Restricciones en la venta de elementos de bioseguridad	No	Sí	Sí
	Prohibición general de anuncios con palabras relacionadas con el covid-19	Sí, con aclaraciones	Sí	N/A
	Desmonetización de contenidos enfocados en el covid-19	No	No	No