
Las medidas de YouTube y TikTok contra la desinformación climática: cómo arreglar un sistema opaco que no funciona

La desinformación sobre el cambio climático puede tener consecuencias extremadamente perjudiciales: afecta negativamente a la preparación ante catástrofes naturales y puede socavar innecesariamente la confianza del público en las órdenes de evacuación y otros avisos de los servicios públicos durante fenómenos meteorológicos extremos. También puede impedir la adopción de políticas públicas eficaces para hacer frente al cambio climático, al presentarlo como algo inocuo o ajeno a la actividad humana, cuando no se llega directamente a negar su existencia.

Estos peligros encajan en algunas de las categorías de riesgos sistémicos contempladas en la Ley de Servicios Digitales (DSA) de la UE, concretamente aquellas que tienen efectos negativos «reales o previsibles» sobre “el discurso cívico y los procesos electorales”, “la seguridad pública”, “la protección de la salud pública” o “el bienestar físico y mental de las personas”. La misma normativa obliga a las plataformas que superan determinado tamaño (entre ellas YouTube y TikTok) a contar con medidas eficaces para mitigar esos riesgos.

La mayoría de las grandes plataformas digitales reconocen estos riesgos hasta cierto punto y cuentan con disposiciones en sus términos y condiciones para limitar los daños. [En un estudio](#) que acabamos de publicar junto con Science Feedback, analizamos cómo YouTube y TikTok aplican sus políticas sobre desinformación climática en español y francés, y cómo los organismos de resolución extrajudicial de litigios (ODS, por sus siglas en inglés) o los reguladores digitales nacionales respondieron a las reclamaciones sobre las decisiones de moderación de las plataformas cuando estas no se ajustaban a sus propias políticas. Los resultados son llamativos.

YouTube, cuyas normas publicitarias prohíben la monetización de contenidos que contienen desinformación sobre el clima, rara vez aplica esa política, incluso cuando se le informa de que un determinado vídeo monetizado la incumple. TikTok, que en teoría no permite la desinformación climática que contradice el consenso científico, la aplica de manera muy inconsistente. Al presentar las reclamaciones ante el Appeals Centre Europe, un organismo de resolución extrajudicial de litigios, y ante el regulador digital francés Arcom, comprobamos que los tiempos de respuesta eran largos y pudimos documentar las limitaciones del proceso. Pero, por el camino, aprendimos algunas lecciones valiosas.

La moderación de la desinformación climática está mal diseñada, se aplica rara vez y no es adecuada para su propósito

El problema de **TikTok** va más allá de su incapacidad para aplicar sus propias políticas: afecta a las propias normas. La retirada de contenidos no es una medida adecuada frente a la desinformación, salvo que esta suponga un riesgo inmediato para las personas; una intervención más apropiada consiste en etiquetar, reducir la visibilidad y/o desmonetizar los contenidos. Las etiquetas alertan al público sobre desinformación sin infringir la libertad de expresión, mientras que la reducción de visibilidad y la desmonetización disminuyen la difusión de la desinformación o desincentivan su creación. Por el contrario, la eliminación puede percibirse como censura..

Nuestros resultados muestran que estas consideraciones también podrían estar afectando a la aplicación de las políticas de TikTok: incluso en el 20% de los vídeos denunciados que contenían desinformación climática y sobre los que TikTok tomó alguna medida, la plataforma optó por reducir su visibilidad en lugar de eliminarlos. Un análisis de los contenidos moderados indica que TikTok estaba aplicando una norma diferente, relativa a las teorías de la conspiración, en lugar de su principal política sobre desinformación climática, que muchos de esos contenidos infringían y que contempla la eliminación del contenido. Aunque consideramos que esa política es un error, creemos que las plataformas tienen la obligación de aplicar sus políticas de manera fiel, tal como exige el artículo 14 de la DSA.

En el caso de **YouTube**, consideramos positiva su política de no remunerar la creación de contenidos que contienen desinformación climática, aunque sea insuficiente. YouTube debería, sin duda, etiquetar los contenidos para proporcionar a los usuarios información adicional cuando se encuentren ante desinformación. Pero incluso si se adoptara un cambio de este tipo, el hecho de que nuestro análisis haya constatado que YouTube no aplica en absoluto sus políticas actuales sobre esta cuestión nos lleva a pensar que, sean cuales sean las políticas, la falta de aplicación efectiva las vuelve irrelevantes. Al fin y al cabo, incluso después de que nosotros le informáramos, YouTube apenas actuó sobre el 4% de los contenidos infractores denunciados.

La falta de transparencia sobre la aplicación de las normas genera problemas que van más allá de la propia transparencia

Nuestra posición es que la respuesta de las plataformas a la desinformación debe centrarse en ofrecer contexto adicional en lugar de retirar contenidos, pero también que cualquier medida de moderación en este ámbito debe priorizar la transparencia y la verificabilidad de la aplicación de las normas respecto de contenidos que sean legales pero perjudiciales. Dicho esto, cuando se aplican sanciones inferiores a la eliminación y la retirada de contenidos queda, en su mayor parte, descartada (como debería ser), la aplicación de las normas depende de medidas como la reducción de visibilidad y la desmonetización, y esas acciones son difíciles de supervisar. Mientras que la retirada de un vídeo deja un rastro, un vídeo cuya visibilidad ha sido reducida no lo hace. Por ello, existe poca información accesible que permita determinar si las plataformas cumplen realmente sus propias políticas.

Esto dificulta exigir responsabilidades a las plataformas y complica la investigación sobre los riesgos sistémicos contemplados en la DSA, especialmente en una época en la que las grandes empresas tecnológicas han [bloqueado agresivamente el acceso de los investigadores a sus datos](#). En lo que respecta a las políticas pertinentes de YouTube y TikTok sobre desinformación climática, que establecen sanciones inferiores a la eliminación, estas no proporcionan datos suficientes para evaluar si sus políticas se están aplicando o no. En estos casos, los investigadores tienen que recurrir a indicadores indirectos, como hicimos nosotros. Por ejemplo: imaginemos tener que revisar más de 80 vídeos para comprobar si, algún tiempo después de haberlos denunciado por infringir las políticas de YouTube, siguen mostrando anuncios antes de su reproducción, lo que indicaría que continúan monetizándose.

Algunas sanciones aplicadas en las plataformas son incluso más difíciles de contabilizar, ya que no existen indicadores indirectos evidentes: puede haber vídeos cuya visibilidad se haya reducido como resultado de una decisión de moderación basada en las políticas sobre desinformación climática, pero incluso el usuario que los denunció inicialmente puede no tener forma de saber si ese es el caso. El artículo 20 de la DSA establece que los denunciantes de una decisión de moderación de una plataforma (por ejemplo, la decisión de reducir o no la visibilidad de un vídeo) tienen derecho a recibir una respuesta a su denuncia. Sin embargo, YouTube, por ejemplo, no ofrece ninguna respuesta en el caso de decisiones que no impliquen el borrado.

Aunque los informes de transparencia de TikTok y YouTube en el marco de la DSA ofrecen cifras agregadas, estas no resultan útiles para realizar una comprobación independiente real de la aplicación de sus políticas. [TikTok](#) afirma que el 64% de los contenidos que modera en virtud de su política sobre desinformación fueron eliminados y que la visibilidad del 36% fue restringida, mientras que [YouTube](#) registró más de 153.000 medidas de moderación en virtud de su política publicitaria, ninguna de ellas consistente en una acción visible. Estas cifras no permiten evaluar la exactitud de las medidas de aplicación ni sus efectos sobre la exposición de los usuarios a este tipo de desinformación.

Los ODS necesitan más recursos y «buena fe» por parte de las plataformas organismos de resolución extrajudicial de litigios

Solo el 10% de las reclamaciones contra decisiones de moderación (en este caso, por moderación insuficiente) que presentamos ante el Appeals Centre Europe se resolvieron en el plazo de un mes. Este periodo de espera resulta inadecuado para abordar eficazmente los daños causados por los contenidos, incluso si las plataformas estuvieran dispuestas a rectificar como consecuencia de la decisión del organismo de resolución extrajudicial de litigio. La responsabilidad por esta demora no recae necesariamente en los ODS, ya que estos han descrito en numerosas ocasiones las limitaciones que encuentran por parte de las plataformas para acceder a la información que necesitan para resolver las reclamaciones.

La normativa no hace vinculantes las decisiones de los ODS, pero exige a las plataformas que cooperen con ellos «de buena fe» y señala la aplicación de las decisiones de los ODS como una de las formas de mitigar los riesgos sistémicos. Ahora bien, si los propios ODS permanecen a oscuras respecto al impacto real de sus resoluciones, esto no solo aumenta la opacidad en torno a las decisiones de moderación de contenidos que se encuentran por debajo del umbral de retirada (y que se toman como consecuencia de las denuncias de los usuarios), sino que también dificulta cualquier evaluación independiente de esa herramienta concreta de mitigación de riesgos.

Las plataformas deben proporcionar **datos sobre las reclamaciones individuales** y los reguladores deben evaluar continuamente si la relación de las plataformas con estos organismos es coherente con el principio de «buena fe» establecido en la normativa. Aunque las decisiones no sean vinculantes, debería esperarse que las plataformas expliquen por qué no las aplican y, si los datos indican que lo hacen a gran escala, esto debería incorporarse a sus evaluaciones de cumplimiento. Como [Husovec señala acertadamente](#), «el efecto de las decisiones de los ODS se describe mejor como “vinculantes sistémicamente, pero no individualmente”». Esto implica que, aunque las plataformas puedan rechazar solicitudes concretas, no deberían poder ignorar sistemáticamente el mecanismo en su conjunto sin ofrecer una justificación.

También debería exigirse una mayor transparencia en materia de monetización, poniendo a disposición información sobre qué cuentas participan en programas de reparto de ingresos en todas las plataformas y sobre cómo se distribuyen esos ingresos entre los creadores, creando vías específicas para la desmonetización dentro de los mecanismos internos de gestión de reclamaciones e informando a los usuarios cuando una denuncia dé lugar a la desmonetización del contenido infractor.

[Accede al informe completo de Fundación Maldita.es y Science Feedback.](#)