

Brasil: Un experimento de Amnistía Internacional revela que Meta no detecta anuncios con desinformación electoral en Facebook

- *En vísperas de las elecciones presidenciales de Brasil, Amnistía Internacional envía anuncios a Facebook para poner a prueba la solidez del sistema de moderación de Meta*
- *Más de la mitad de los anuncios que difunden desinformación relacionada con las elecciones son aceptados*
- *Los demás se identifican, aunque no son rechazados expresamente por contener desinformación electoral*

Amnistía Internacional ha advertido hoy que Meta podría no estar abordando de manera eficaz la difusión de desinformación electoral a través de la plataforma publicitaria de Facebook en vísperas de las elecciones presidenciales brasileñas del 4 de octubre.

Las redes sociales llevan años usándose para difundir desinformación en Brasil. Con objeto de investigar cómo gestiona Facebook en su publicidad la desinformación relacionada con elecciones, Amnistía Internacional llevó a cabo un experimento consistente en presentar en esa red social anuncios “falsos” que contenían ese tipo de desinformación. El contenido de los anuncios iba desde denunciar fraudes electorales a utilizar la desinformación para promover un golpe de Estado militar. El resultado del experimento fue que ocho de los anuncios presentados fueron aceptados, mientras

que los siete restantes se marcaron para verificar la identidad del anunciante, por considerarse relacionados con temas sociales, elecciones o política, pero no fueron rechazados específicamente por contener desinformación electoral. Pese a que Facebook no rechazó los anuncios de prueba, Amnistía se aseguró de que no llegaran a publicarse.

“Resulta sumamente alarmante que el sistema de moderación de Meta no haya detectado más de la mitad de los anuncios relacionados con las elecciones que contenían desinformación. Demuestra que Meta no está haciendo lo suficiente para evitar que su plataforma sea un vector de desinformación, mientras sigue obteniendo beneficios de los contenidos y los sistemas de publicidad que hacen posible su expansión”, ha declarado Jurema Werneck, directora ejecutiva de Amnistía Internacional Brasil.

El experimento

Amnistía Internacional presentó a Facebook 15 anuncios en portugués de Brasil, dirigidos exclusivamente a la población brasileña en edad de votar, para evaluar cómo los sistemas de moderación publicitaria de la plataforma trataban los contenidos relacionados con las elecciones en vísperas de la cita electoral del 4 de octubre de 2026.

Para que los anuncios se basaran en narrativas electorales reales que circulaban en Brasil, la organización analizó conversaciones online y la opinión pública en redes sociales, así como contenidos en portugués publicados en las plataformas Facebook y X entre los meses de enero y mayo de 2026.

De los 15 anuncios, todos contenían desinformación menos uno, que era neutro. Este análisis utiliza el término “desinformación” para mantener la coherencia con la terminología de Amnistía Internacional, pero en el experimento se aplica la definición de Meta de “información errónea” en la construcción y el análisis de los anuncios. Como indican sus

políticas y sus Normas Comunitarias, esta definición abarca información falsa o inexacta, incluidos contenidos que es probable que contribuyan a interferir con el funcionamiento de procesos políticos.

Los 14 anuncios con desinformación fueron concebidos para que reflejaran tres tácticas utilizadas a menudo por políticos/as y candidatos/as para apoyar prácticas autoritarias: “la deslegitimación electoral” (contenidos cuya intención es socavar la confianza en las elecciones, los sistemas de votación o los resultados electorales); “la construcción de un enemigo” (presentar a los oponentes políticos y los tribunales como amenazas a la nación o actores ilegítimos); y “el autoritarismo centrado en la seguridad” (presentar la fuerza, la intervención del ejército o el autoritarismo de mano dura como respuestas necesarias a la delincuencia o el desorden político).

Algunos anuncios se crearon en torno a una única categoría, pero otros incorporaban narrativas de varias. Todos los anuncios, excepto el neutro, violaban la política de Meta sobre información errónea, en concreto su apartado “Interferencia electoral o censal”. El anuncio neutro se incluyó para poner a prueba el requisito de Meta de verificar la identidad para todos los contenidos relacionados con procesos electorales.

Siete de los anuncios creados por Amnistía Internacional se centraron en afirmaciones deliberadamente engañosas sobre los candidatos y las instituciones electorales, por ejemplo las acusaciones de que ciertos candidatos estaban amañando las elecciones, y cuestionando la integridad del sistema electoral. Los otros siete incluían información falsa que podía afectar a la capacidad de las personas para votar, por ejemplo detalles incorrectos sobre cuándo o cómo hacerlo, o sobre quiénes se presentaban a las elecciones.

En la cuenta de Facebook que se utilizó para presentar los

anuncios no se llevó a cabo el proceso de Meta de verificación de la identidad ("Confirmar tu identidad"). La verificación de la identidad no se llevó a cabo como parte de la investigación: así se mantenía una cuenta de prueba neutra, y se evitaba que el historial o la actividad previa de la cuenta influyeran en los resultados de la moderación de anuncios. Se utilizaron cuentas exclusivas para la investigación, no cuentas personales, para reducir el riesgo de que el historial de uso de la cuenta influyera en los resultados.

Quienes realizaron la investigación no se encontraban físicamente en Brasil, y no utilizaron una VPN para que pareciera que los anuncios se presentaban desde Brasil (se presentaron desde Londres).

La publicación de todos los anuncios se fijó para una fecha futura. De esta manera, Amnistía Internacional podía poner a prueba los procesos de revisión de las plataformas y al mismo tiempo se garantizaba que los anuncios aceptados pudieran cancelarse con antelación a esa fecha, impidiendo así su publicación y evitando posibles daños para los derechos humanos u otros riesgos de tipo ético.

Amnistía Internacional se dirigió por escrito a Meta el 21 de julio de 2026, solicitándole que respondiera a las acusaciones de que los sistemas de moderación de anuncios de Facebook no habían detectado desinformación relacionada con las elecciones y expresándole su preocupación respecto a la idoneidad de los sistemas de la empresa para identificar, evaluar y mitigar riesgos relacionados con los derechos humanos. En el momento de redactar estas líneas no se ha recibido respuesta de Meta.

Conclusiones

Facebook aprobó ocho anuncios que contenían desinformación, y ninguno de los otros siete fue rechazado explícitamente por contener desinformación.

De los ocho anuncios que contenían desinformación y fueron

aceptados para su publicación, cuatro se centraban únicamente en la “deslegitimación electoral”, dos pertenecían únicamente a la categoría de “autoritarismo centrado en la seguridad”, e incluían desinformación que abogaba por la intervención militar o la abstención electoral, y el contenido de los otros dos se solapaba entre las tres categorías.

Los siete anuncios rechazados, incluido uno que era neutro, lo fueron por tratarse de “anuncios sobre temas sociales, elecciones o política”. En estos casos, el Administrador de Anuncios de Facebook señaló que, dado que los anuncios parecían hacer referencia a temas sociales, elecciones o política, era preciso completar la verificación de identidad de la cuenta para que pudieran publicarse. Es importante señalar que estos avisos de rechazo no indicaban que los anuncios hubieran sido rechazados por su contenido, y, por consiguiente, no proporcionaban indicación alguna sobre cómo habrían sido valorados si la cuenta hubiera sido verificada.

Amnistía Internacional también detectó deficiencias en las salvaguardias aplicadas en la presentación de anuncios relacionados con las elecciones. Se permitió su presentación sin declarar que estaban incluidos en las Categorías de asuntos especiales de Facebook, como exige la ley electoral de Brasil, así como las propias políticas de Meta relativas a los anuncios relacionados con las elecciones. Como la identidad de la cuenta no estaba verificada, también se permitió que los anuncios se presentaran sin el descargo de responsabilidad con la etiqueta “Pagado por”, como exigía Meta para contenidos relacionados con las elecciones. Se trata de anuncios que, para empezar, Facebook no debería haber permitido que se presentaran sin verificación de identidad.

El hecho de que los anuncios pudieran ser presentados desde fuera de Brasil plantea la preocupación de que unos controles inadecuados de verificación e información puedan crear vulnerabilidades que podrían ser aprovechadas por actores nacionales o extranjeros para tratar de influir en los

procesos electorales. Este incumplimiento de la obligación legal de publicar la información electoral requerida, y la presentación de los anuncios desde fuera de Brasil, plantean serias preocupaciones sobre el riesgo de que haya campañas de desinformación, tanto por parte de agentes nacionales como internacionales, que interfieran en los procesos electorales, además de suscitar interrogantes más amplios sobre los sistemas de moderación de Meta para identificar estos peligros.

El sistema de aplicación de las políticas de Meta utiliza tecnología de inteligencia artificial para identificar proactivamente el contenido de anuncios que podrían violar sus políticas. A continuación, dicho contenido es evaluado, bien sea automáticamente, o bien por personas que lo revisan y, cuando se confirma una violación, se llevan a cabo acciones de aplicación, como la retirada, la restricción o el rechazo de los anuncios con arreglo a las Normas comunitarias de Meta.

“La investigación indica que los sistemas de moderación de Facebook podrían no estar detectando de forma fiable contenido con desinformación relacionado con las elecciones. Si bien Facebook rechazó siete de los 15 anuncios, los marcó con arreglo a su política sobre temas sociales, elecciones o política, no específicamente por desinformación electoral. Esto suscita honda preocupación respecto a si se están aplicando de forma coherente las salvaguardias sobre temas electorales”, ha declarado Ummar Bashir, investigador y asesor sobre Inteligencia Artificial y Derechos Humanos.

Amnistía Internacional halló que los sistemas de moderación de Facebook parecían más eficaces en el reconocimiento de anuncios que presentaban a oponentes políticos como “delincuentes”, “traidores” o “agentes ilegítimos”. Por ejemplo, uno de los anuncios de prueba afirmaba que “Lula [el presidente de Brasil] robará las elecciones”. Cinco de cada siete anuncios con este tipo de contenidos fueron rechazados. Por el contrario, tres de cada cuatro anuncios que contenían

retórica militar dirigida contra el Estado de derecho y pedían que el ejército tomara el poder, superaron los sistemas de moderación de Facebook. En uno de ellos se afirmaba: “¡Necesitamos una revolución militar, no voten en las elecciones, el ejército recuperará el poder si la abstención supera el 50%!”

“En vez de salvaguardar la participación electoral, el acceso a información veraz y los derechos humanos, con estos sistemas hay peligro de que se apruebe contenido dañino”, ha afirmado Ummar Bashir, investigador y asesor sobre Inteligencia Artificial y Derechos Humanos.

Normas sobre empresas y derechos humanos

Según las normas internacionales, las empresas tienen la responsabilidad de respetar los derechos humanos y evitar causar, contribuir o estar directamente vinculadas a impactos adversos sobre los derechos humanos. Esto incluye tomar medidas proactivas para identificar, prevenir y mitigar los impactos sobre los derechos humanos.

En el caso de plataformas como Meta, supone garantizar que las prácticas de moderación no permitan que circulen contenidos falsos o inexactos que puedan afectar al derecho de las personas a la información y a participar en asuntos públicos. La responsabilidad de respetar los derechos humanos existe con independencia de la normativa del Estado y exige una constante evaluación de riesgos, así como transparencia y medidas de mitigación eficaces.

El modelo de negocio basado en la vigilancia de Meta se basa en la recopilación y el análisis a gran escala de datos de usuario para mantener a la gente activa en sus plataformas y hacer posible una publicidad muy segmentada. Estos anuncios se muestran mediante una compleja combinación de las características del perfil de usuario, que incluye ubicación, datos demográficos, intereses y comportamiento online. La

capacidad de Meta de ofrecer a los anunciantes herramientas predictivas y de microsegmentación de gran precisión es fundamental para este modelo. En el primer trimestre de 2026, el 98% de los ingresos de Meta procedía de la publicidad.

En vísperas de las elecciones presidenciales de Brasil, Amnistía Internacional pide a Meta que garantice que se dispone de recursos lingüísticos adecuados en portugués brasileño para la moderación de contenidos, incluida la aprobación de anuncios, y que refuerce con carácter inmediato su capacidad de identificar y responder a peligros relacionados con las elecciones.

Meta también debe llevar a cabo y publicar un análisis de diligencia debida en materia de derechos humanos, actualizado y específico para cada país, que evalúe cómo sus sistemas, incluida la moderación de contenidos y la publicidad, pueden contribuir a generar riesgos para la participación y el acceso a la información, y adoptar medidas eficaces y transparentes para mitigar estos riesgos.

“A poco más de cinco semanas de las elecciones, Meta aún está a tiempo de abordar con urgencia los riesgos que presentan sus sistemas de moderación de anuncios y evitar el uso de su plataforma para difundir desinformación peligrosa. Permitir que semejante contenido circule sin control podría confundir a los votantes, minar la confianza en los procesos electorales y poner en riesgo la capacidad de la gente para ejercer de manera libre y significativa su derecho al voto”, ha afirmado Jurema Werneck, directora ejecutiva de Amnistía Internacional Brasil.

Información complementaria

En los últimos años, Brasil ha implementado varias medidas para abordar la desinformación relacionada con las elecciones. Durante las elecciones presidenciales de 2022, la autoridad electoral de Brasil, el Tribunal Superior

Electoral, introdujo medidas más estrictas, que exigían que las plataformas online eliminaran rápidamente la información falsa o deliberadamente descontextualizada sobre los procesos electorales, so pena de ser multadas y de su posible suspensión.

El uso de inteligencia artificial (IA) también se restringió antes de las elecciones municipales de 2024 para evitar la desinformación y la manipulación del proceso electoral. En 2024, el Tribunal Supremo de Brasil ordenó también la suspensión de X (antes Twitter) por incumplimiento de órdenes judiciales, entre ellas algunas relacionadas con cuentas acusadas de difundir desinformación.

En marzo de 2026, una nueva resolución señaló los contenidos específicos que las plataformas deben eliminar, incluido contenido falso o descontextualizado sobre el sistema de votación electrónica, el Tribunal Electoral y otros procesos electorales, y también prohibió el uso de IA generativa para clasificar, recomendar, sugerir o priorizar candidaturas, campañas y partidos políticos.