«Existe una red que disemina la desinformación desde distintos países, pero con una identidad ideológica común»

Por: Álex Blasco Gamero. 20/01/2025

Elías Said-Hung (Valencia, Venezuela, 1979) es catedrático e investigador en la Universidad Internacional de La Rioja. Durante años ha estudiado en profundidad los procesos sociales aplicados en las redes sociales, en especial en temas relacionados con la desinformación, la participación ciudadana y las expresiones de odio. Actualmente codirige Hatemedia, un ambicioso proyecto de monitoreo de los mensajes de odio en las plataformas digitales de cinco de los principales medios digitales de España. Tras el abandono de toda promesa de equidad y diversidad y de la lucha contra los mensajes de odio y la desinformación por las grandes tecnológicas de los últimos días es el momento perfecto para intentar entender estos movimientos.

¿En qué consiste el monitoreo que han realizado sobre el discurso de odio en los medios digitales?

Nuestro objeto de estudio son los espacios de comunicación de los medios digitales en España. Seleccionamos los cinco principales medios digitales según el número de usuarios/lectores en 2020 (*La Vanguardia, ABC, El País, El Mundo y 20 Minutos*). A partir de ahí, recopilamos los contenidos publicados y de forma segregada los comentarios recibidos en redes sociales y los comentarios en la propia web. Este enfoque nos permite extraer y categorizar los mensajes para identificar niveles de discurso de odio y estudiar su impacto en la opinión pública.

Algunos resultados son llamativos. Por ejemplo, el 38% de los comentarios analizados en redes sociales son clasificados como odio. ¿Qué explica este porcentaje tan alto?

Ese 38% incluye mayoritariamente mensajes de odio de baja intensidad, como mensajes incívicos. Aunque estos no impliquen amenazas directas, contribuyen a la

hostilidad y la polarización del debate público. Es crucial entender que el odio de baja intensidad puede ser igual de perjudicial a largo plazo, ya que crea un entorno tóxico en los espacios digitales. Además, hemos observado que este tipo de mensajes muchas veces son ignorados o incluso normalizados, lo que refuerza su presencia en los debates. El odio ha sido estudiado desde el simple lenguaje ofensivo a comportamientos incívicos o violentos. ¿Existen muchas amenazas violentas denunciables según lo tipificado en la ley? Sí, pero lo que más hay es odio de baja intensidad. Un comportamiento que normaliza la hostilidad entre usuarios y la polarización de la opinión pública.

En una presentación reciente mencionó que el odio en redes no es del todo espontáneo, sino que sigue ciertos patrones. ¿Podría explicar esto?

En enero de 2021 recogimos una muestra de mil usuarios para analizar. En ella observamos que muchos de esos usuarios tenían una correlación directa en cuanto al uso de temas específicos y que a nivel de plataforma se repetían patrones de escritura entre personas/usuarios. Ya sea en textos cortos o largos, todos repetimos unos patrones personales. Esto sugiere una organización, y que en muchas ocasiones una misma persona está detrás de varios perfiles. Hablamos de usuarios especializados en plataformas concretas que actúan de forma sistematizada según el momento, lo que les permite maximizar su impacto. Esta organización no es casual. Estos resultados están respaldados por investigaciones internacionales del Departamento de Seguridad Nacional de EEUU y de la UE, que muestran cómo existen granjas de *trolls* o de *bots* que diseminan mensajes de manera coordinada para influir en la opinión pública. Es muy difícil que cincuenta parejas de baile en un salón hagan los mismos pasos de forma sincronizada, algo que estamos viendo a diario con noticias relacionadas con la inmigración o movimientos sociales en las redes sociales.

¿Qué rol juegan las granjas de *trolls* y *bots* en este fenómeno?

Son fundamentales. En las primeras etapas del estudio de las redes sociales se hablaba mucho de *bots*. Sin embargo, las granjas de *trolls* han ganadoprotagonismo. Estas son personas especializadas en una plataforma concreta conuna cartera multicuenta de usuarios falsos que publican mensajes orgánicos usandola estrategia *astroturfing*.

¿Por qué es esto? Muy sencillo. Entre pagar a un ingeniero que esté actualizando constantemente los algoritmos de emisión de los mensajes de los *bots*, que además son más fáciles de detectar, y pagar una granja de *trolls* suele ser más rentable lo segundo. Estas granjas operan en países como Filipinas, Venezuela, México, Rusia o Estados Unidos, donde los costes de personal suelen ser bajos.

Mencionó el concepto de *astroturfing*. ¿Podría explicarlo?

El *astroturfing* es una estrategia de desinformación que se remonta a la Segunda Guerra Mundial. Esta fue creada por el Ejército aliado para contrarrestar la propaganda nazi y, principalmente, consiste en crear la apariencia de un apoyo popular espontáneo a través de mensajes coordinados. Esto suele implicar usuarios alfa, que lanzan los mensajes iniciales, y usuarios beta, que los refuerzan con comentarios. Muchos de estos perfiles son *nanoinfluencers*, perfiles con pocos seguidores, poca interacción y que llevan poco tiempo en la plataforma, lo que los hace difíciles de detectar. Este tipo de estrategias buscan influir en la opinión pública de manera sutil pero efectiva. Por ejemplo, un mensaje inicial puede sembrar una idea polémica y luego ser amplificado mediante una serie de respuestas masivas para dar la impresión de que representa una opinión mayoritaria. ¿A qué nos lleva esto? A la instauración de una promoción de un poder suave.

¿Tiene algún objetivo concreto este poder suave?

Este poder de influencia está articulado por conveniencia. En este momento su objetivo principal es la inmigración, el movimiento *woke*, intereses concretos en relacionar comportamientos ‘incívicos’ con colectivos minoritarios y en la promoción de teorías conspiranoicas, como es la del ‘gran reemplazo’.

Lo peligroso de este poder suave coordinado es el impacto que trae en la cultura democrática, basada en el diálogo y el respeto. Nuestra sociedad se ha

sentimentalizado, adquiriendo mecanismos de comunicación propios de las redes sociales, principalmente por la reafirmación de las ideas propias. Si a esto le añadimos una participación de actores concretos que lo que buscan es normalizar altos niveles de tolerancia a la hostilidad y desinformar, lo que conseguiremos es tener una sociedad cada vez más crispada y menos dialogante, menos democrática. Hay que tener en cuenta que el odio siempre ha existido, la clave en este momento es luchar para que este no sea tan apabullante y ocupe todos los espacios.

Por los temas que menciona, los actores que participan en este poder suave parecen tener una identidad clara.

En las plataformas existen mensajes tanto de extrema izquierda como de extrema derecha pero lo que es evidente es su asimetría tanto en volumen como en alcance. Hay un estudio reciente de la revista *Science* que habla de los dilemas asociados a la desinformación y el odio. Analizan su presencia en las elecciones al Parlamento Europeo en Italia de 2019, las elecciones presidenciales de EEUU de 2020 y las de Filipinas de 2016. En todos los casos encuentran un patrón común: la mayoría de los bulos y mensajes de odio detectados en redes sociales eran identificados como de extrema derecha, eran consumidos por usuarios de derechas y/o favorecían a algún candidato conservador. Algo que nosotros también hemos visto aquí.

El principal problema de esto es que estamos tratando de analizar un fenómeno global desde perspectivas concretas cuando es evidente que las estrategias son internacionales. Existe una red que disemina esta información desde distintos países a nivel mundial con especialistas en plataformas muy diferentes, pero con una identidad ideológica común.

Según su estudio, ¿qué tipo de odio predomina en las redes sociales?

En general, el discurso de odio de baja intensidad es el más frecuente. Representa aproximadamente el 70% de los mensajes que analizamos. Este tipo de odio no suele estar tipificado como delito, pero contribuye a un ambiente hostil y polarizado. En cuanto a temáticas, el odio político es el más recurrente, seguido por la xenofobia y la misoginia. Sin embargo, esto varía según la actualidad informativa y los eventos que marquen la agenda mediática. Lo que sí hemos detectado es que los periodistas y los políticos se están convirtiendo en los principales receptores de ataques. Estás

acciones no son solamente por la vinculación con el medio o partido sino por lo que representan ideológicamente o por pertenecer a un colectivo vulnerable. El proceso de ataque de odio utiliza al protagonista de la noticia como mecanismo de desvío del tiro hacia los rasgos o características que comparte con un grupo en específico.

En los últimos días Meta ha decido prescindir de los verificadores de datos y eliminar sus programas de equidad, diversidad e inclusión de su reglamento. ¿Hasta qué punto las plataformas avivan más este modelo?

Las plataformas deberían tener la responsabilidad de mejorar sus mecanismos de moderación, pero tenemos que tener muy clara su finalidad: cuanto más debate y más interacción, más datos y por lo tanto mayor beneficio económico. Con la compra de Twitter por Musk la plataforma ha ido a peor, pero la tendencia previa ya era negativa. Y no hay que olvidar el escándalo de Cambridge Analytica con Facebook. Son ya varios los ejemplos. No podemos dejar al lobo cuidar de las ovejas y pensar que va actuar en beneficio de estas.

Es crucial que las plataformas inviertan en tecnología que permita detectar patrones de odio de forma más eficiente y que los gobiernos establezcan marcos regulatorios claros. El Centro de Excelencia de Comunicaciones Estratégicas de la OTAN realizó entre agosto y septiembre de 2024 un experimento muy sencillo: crearon un *bot* con 58 euros para diseminar bulos y expresiones de odio en Facebook, Instagram, YouTube, TikTok, VKontakte y Twitter/X, esperaron un tiempo y denunciaron esos mismos mensajes en cada plataforma. Después de tres meses, el resultado de mensajes eliminados fue tan solo del 6 %.

¿Las instituciones pueden o, mejor dicho, están haciendo algo?

Ahora mismo la UE está avanzando en el diseño, basado en el modelo *DISARM,* de estrategias de ataque y defensa contra la desinformación a raíz de la guerra de Ucrania. Este es un modelo que ya se aplica en la ciberseguridad de los sistemas bancarios y que ayuda a los equipos a comprender, detectar y prevenir las ciberamenazas persistentes. En el momento que esto se desarrolle se podría trasladar a la lucha contra el odio.

Si dejamos a un lado el abandono total de las redes sociales, ¿qué podemos hacer como usuarios?

Los usuarios pueden contribuir de forma reactiva, denunciando mensajes o participando en contranarrativas, o de forma proactiva, planteando unas normas de participación o bloqueando usuarios. Aunque entendemos que esto puede ser complicado debido a las represalias; no hacerlo implica ceder ante el que más grita, el que más acusa y el que más insulta. En una submuestra de un mes analizamos la figura de la mujer política. El porcentaje de mensajes de odio misógino que detectamos fue de más de un 40%. La mayoría concentrado en odio incívico, malintencionado y, sobre todo, insultante. En donde se cosificaba a la mujer, se la desprestigiaba atacando sus capacidades personales y se la relacionaba con hechos de corrupción o criminalidad. El porcentaje de comentarios publicados que rebatían esos mensajes de odio fue de menos del 16%. El 40% de los mensajes restantes fueron de usuarios que, viendo ese comportamiento, prefirieron no hacer nada.

El próximo día 20 se ha programado una gran huida de Twitter/X, por ser un entorno demasiado nocivo. ¿Cuánta culpa tiene la plataforma y cuánta los comportamientos que hemos ido normalizando en ella?

La alarma de un coche no evita que nos puedan robar el coche, pero les complica el trabajo a los ladrones. Si partimos de ese principio, lo ideal sería exigir un marco jurídico que luche contra el odio y un marco empresarial que persiga las expresiones de odio y que desarrolle tecnología para cancelar granjas de *bots* o de *trolls*.

Por ejemplo, Bluesky ahora tiene rasgos comunes con los del comienzo de Twitter, pero es erróneo pensar que las redes sociales son espacios felices por naturaleza. En este caso tienen herramientas para detectar y rechazar mensajes de odio, pero su algoritmo se tiene que adaptar, en nuestro caso, a los distintos tipos de español, aplicar un proceso de detección permanente y evolucionar a la vez que aparecen nuevas formas de odio. Esto implica un desarrollo técnico continuo, y unagrandísima inversión. Por lo tanto, todo dependerá de los marcos legales, del interésempresarial y, en último lugar, del comportamiento de los usuarios. No hay queolvidar que los odiadores también se terminarán mudando.

LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ

Fotografía: CTXT. Elías Said-Hung

Fecha de creación
2025/01/20