



TRABAJOS DEL FORO CONTRA LAS CAMPAÑAS DE DESINFORMACIÓN

INICIATIVAS 2024

CAPÍTULO 5

CAMPAÑAS DE DESINFORMACIÓN Y PROMOCIÓN DEL DISCURSO DE ODIO

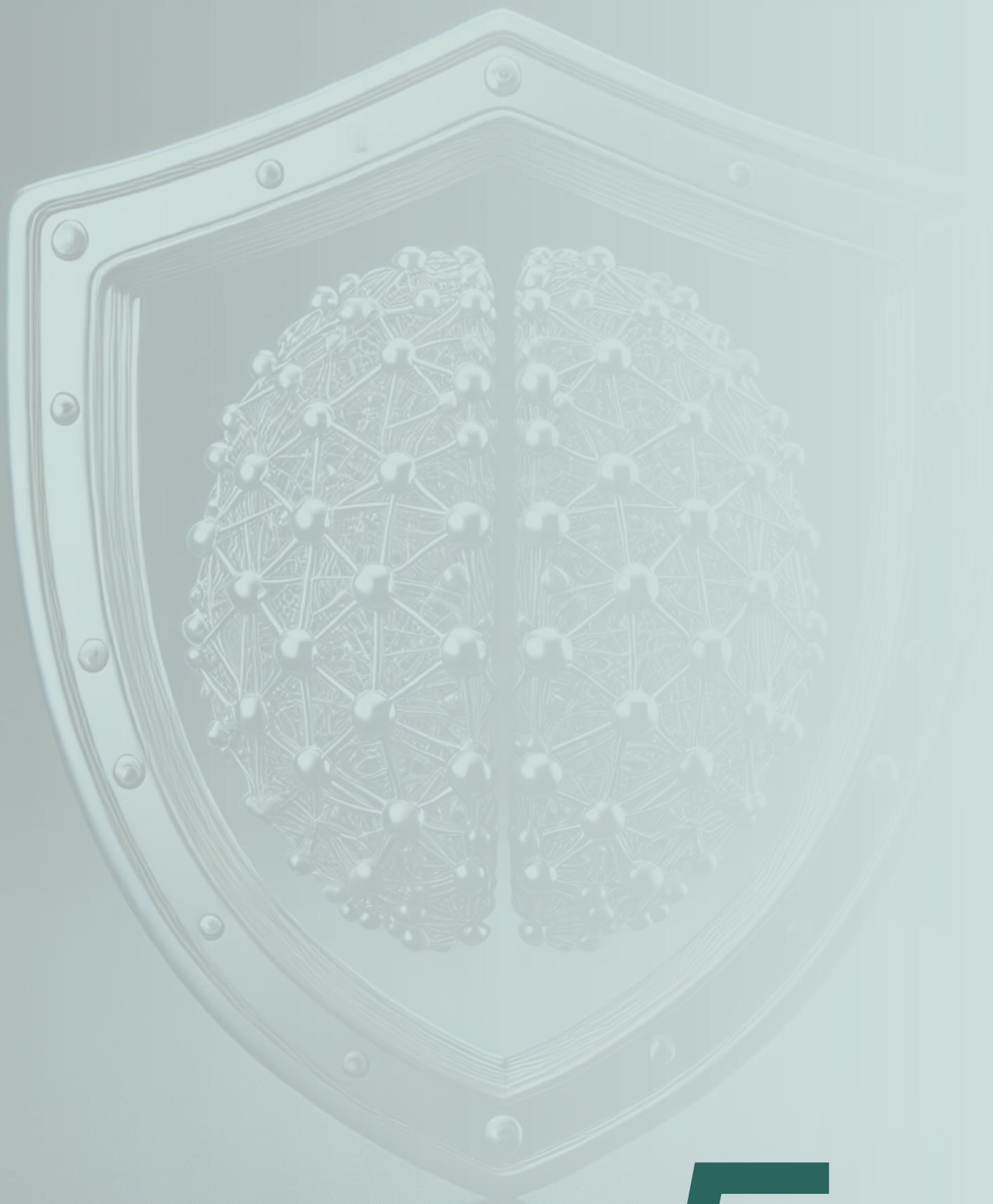
Los expertos participantes en los Grupos de Trabajo lo son a título personal y no a título institucional. Por lo tanto, sus opiniones y recomendaciones no representan ni comprometen a las instituciones a las que pertenecen.

El resultado de los trabajos es producto de un ejercicio de reflexión colectivo, si bien, no tiene por qué representar la opinión individual de todos los participantes, quienes no necesariamente comparten todas las conclusiones o propuestas.

ÍNDICE

CAPÍTULO 5	4
CAMPAÑAS DE DESINFORMACIÓN Y PROMOCIÓN DEL DISCURSO DE ODIO	5
INTRODUCCIÓN.....	7
DESARROLLO DE LA INICIATIVA.....	8
CONCLUSIONES.....	9
Conceptualización de la amenaza	9
Marco legal y normativo	10
Ámbito tecnológico	13
Alfabetización mediática, concienciación y el papel del tercer sector	14
RECOMENDACIONES.....	16
REFERENCIAS BIBLIOGRÁFICAS	19





CAPÍTULO 5

CAMPAÑAS DE DESINFORMACIÓN Y PROMOCIÓN DEL DISCURSO DE ODIO

Coordinadores:

Mario Hernández Ramos

Miguel Camacho Collados

Departamento de Seguridad Nacional

Autores y colaboradores:

Rubén Arcos Martín

Jesús Díaz Carazo

Carlos Edmundo Arcila

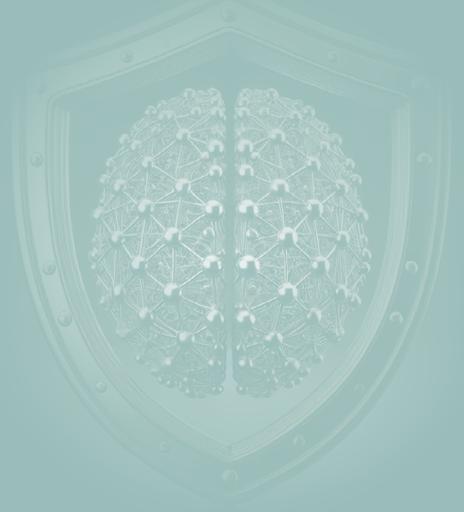
Carmen Girón Tomás

Beatriz Marín García

María Teresa Martín Valdivia

Unidad de delitos de odio y discriminación de la Fiscalía General del Estado

Unidad Especializada contra la Criminalidad Informática de la Fiscalía General del Estado



INTRODUCCIÓN

El riesgo que las campañas de desinformación suponen para alentar y promover el discurso de odio contra determinados colectivos en el seno de los países democráticos ha sido ya foco de atención del Parlamento Europeo (Szakács & Bognár, 2021), el Servicio Europeo de Acción Exterior (EEAS Stratcom Division, 2023) y el Centro Europeo de Excelencia para la Lucha contra las Amenazas Híbridas (Hoogensen Gjørv & Jalonen, 2023), especialmente cuando estas campañas son promovidas por actores estatales hostiles que buscan generar división social. Además, estas campañas también se despliegan en terceros países (principalmente de África y Europa del Este) como una forma de atacar a Occidente y el modelo de democracia liberal, con el consiguiente riesgo que esto genera en los colectivos de estos países.

A esto se le une, la creciente preocupación por el potencial que tiene la desinformación para promover la radicalización violenta, especialmente cuando adopta la forma de teoría de la conspiración que presenta a ciertos grupos o colectivos como una amenaza (Red de Concienciación sobre la Radicalización (RAN), 2020).

Por ello, el 29 de febrero de 2024 el Foro contra las campañas de desinformación en el ámbito de la seguridad nacional acordó la creación de un grupo de trabajo con el objetivo principal de analizar el riesgo que las campañas pueden suponer para el fomento del discurso de odio, con especial atención a España, y discutir los posibles instrumentos o iniciativas para limitar su impacto en nuestra sociedad. Asimismo, el grupo tendría como objetivo fomentar la concienciación y el conocimiento sobre esta amenaza entre los diferentes actores de la sociedad civil y la administración pública implicados en la protección de los colectivos que puedan ser objetivo de estas campañas de desinformación.

Para coordinar este grupo de trabajo se designó a Mario Hernández Ramos y a Miguel Camacho Collados, así como al Departamento de Seguridad Nacional, por parte de la administración pública, como coordinadores. El grupo de trabajo se conformó con las personas que figuran al inicio de este capítulo.

DESARROLLO DE LA INICIATIVA

A fin de alcanzar el objetivo de conocer mejor la amenaza y el marco actual para abordarla, así como para fomentar la concienciación sobre la misma en el seno de las comunidades vinculadas tanto a la lucha contra las campañas de desinformación como a la lucha contra el discurso de odio, se consideró la realización de una jornada con expertos en la materia.

Esta jornada se celebró el 18 de septiembre en la Facultad de Derecho de la Universidad Complutense de Madrid y contó con la colaboración del Instituto de Derecho Parlamentario en el marco del proyecto de investigación "Fortalecimiento de la democracia y el Estado de Derecho a través de la inteligencia artificial"¹. La sesión estuvo organizada en cuatro paneles, dirigidos a abordar diferentes aspectos de la amenaza o de las posibles herramientas disponibles para limitar sus efectos.

El primer panel, orientado a conocer mejor esta amenaza, las técnicas empleadas y cómo se vinculan en la actualidad los fenómenos de las campañas de desinformación y el discurso de odio, estuvo formado por Beatriz Marín, del Servicio Europeo de Acción Exterior; Ruben Arcos Martín, profesor contratado Doctor de la Universidad Rey Juan Carlos; Alicia Moreno Delgado, profesora de la Universidad Internacional de La Rioja; Raquel Godos, de EFE Verifica; y Alejandro González, del Departamento de Seguridad Nacional, como moderador.

El segundo panel estuvo dedicado a conocer las actuales herramientas normativas y legales con las que cuenta España frente a estas amenazas y los retos y oportunidades de nuevos instrumentos como la Ley de Servicios Digitales. La mesa contó con Miguel Ángel Aguilar, Fiscal de Sala Coordinador de la Unidad de los Delitos de Odio y Discriminación de la Fiscalía General del Estado; Karoline Fernández de la Hoz Zeitler, Directora del Observatorio Español del Racismo y la Xenofobia del Ministerio de Inclusión, Seguridad Social y Migraciones; Carlos Aguilar Paredes, de la Comisión Nacional de los Mercados y la Competencia; y Alfonso Peralta Gutiérrez, Juez de Primera Instancia e Instrucción; y estuvo moderada por Rafael Bustos Gisbert, Catedrático de Derecho Constitucional de la Universidad Complutense de Madrid.

En tercer lugar, se abordaron las oportunidades que la tecnología ofrecía para detectar estas amenazas o para contrarrestarlas desde el campo de la comunicación. El panel estuvo integrado por Emilio Delgado López Cózar, catedrático de la Universidad de Granada; David Blanco Herrero, investigador posdoctoral en la Universidad de Ámsterdam (Países Bajos); Gavin Abercrombie, profesor ayudante en la Universidad Heriot Watt (Reino Unido); Flor Miriam Plaza del Arco, investigadora en la Universidad Bocconi (Italia); y estuvo moderado por Maite Martín Valdivia, catedrática de la Universidad de Jaén y Carlos Arcila Calderón, profesor titular de la Universidad de Salamanca.

El último panel puso el foco en la alfabetización mediática y el papel del tercer sector y reunió a expertos en la materia como Manuel Gértrudix Barrio, catedrático de la Universidad Rey Juan Carlos de Madrid; Alberto Izquierdo Montero, profesor ayudante doctor de la UNED; Pablo Hernández Escayola, Coordinador de Investigación Académica de Maldita.es; Natalia Sancha, del Servicio Europeo de Acción Exterior; Marisa Gómez, directora de la Plataforma de ONG de Acción Social; y Carmen Girón Tomás, doctoranda en Derecho y Ciencias Sociales, UNED, quien coordinó el panel.

La conferencia contó con un público de más de setenta asistentes del ámbito académico, expertos de organizaciones de la sociedad civil y centros de pensamiento, así como representantes de la administración pública.

¹ Proyecto financiado por Ministerio de Ciencia e Innovación con PID2021-122677NB-I00.

CONCLUSIONES

Conceptualización de la amenaza

Las campañas de desinformación son entendidas hoy en día en el ámbito de la seguridad nacional como patrones de comportamiento de carácter manipulativo llevados a cabo de forma coordinada e intencional con el objetivo de menoscabar los principios, valores y procesos democráticos. El carácter manipulativo puede apreciarse en la construcción del mensaje (por ejemplo, mediante el uso de *deepfakes*), en la fuente (por ejemplo, mediante la suplantación de medios de comunicación o cuentas oficiales) o en la distribución (por ejemplo, mediante el uso de cuentas automáticas o *bots*).

Los actores que despliegan este tipo de campañas de desinformación con objetivos geopolíticos suelen utilizar vulnerabilidades existentes en la sociedad, ya sea en el ámbito social, económico, político o histórico. En este sentido, en los últimos años cada vez se ha observado la utilización de la identidad como vector de ataque, entre ellas, las vinculadas a género, orientación sexual, raza, etnia o religión.

Estas campañas a menudo tienen como objetivo el fomento de divisiones sociales, menoscabando la cohesión social, así como el ataque a líderes políticos o sociales contrarios a los intereses de los actores hostiles que despliegan las campañas. Estas, además, pueden disuadir a un grupo social de participar en la vida pública y en el proceso político, lo cual genera un menoscabo del sistema democrático. Este tipo de estrategias deben entenderse como acciones de influencia coercitiva (Alonso-Villota & Arcos, 2024).

Entre las estrategias de las campañas de desinformación que usan la identidad como vector, se identifican: la acusación a una persona de tener esa identidad, el intento de modificar la percepción que se tiene en la sociedad sobre esa identidad, el fomento del acoso en línea o incluso la promoción de acciones de acoso o delitos de odio en el plano físico.

Actualmente, en España, los análisis realizados en el ámbito académico han puesto de manifiesto que los discursos de odio en Internet predominantes son de temática racista. En este sentido, los análisis de los verificadores han constatado que la desinformación asociada al discurso racista o xenófobo es la predominante en el ecosistema español. Por ejemplo, el 20% de la desinformación desmentida por los verificadores durante las elecciones al Parlamento Europeo de 2024 estaba dirigida a vincular criminalidad e inmigración.

Si bien no toda la desinformación identitaria puede ser considerada discurso de odio, uno de los retos a afrontar es conocer en qué grado estas campañas de desinformación pueden fomentar el odio, la discriminación o incluso, la violencia contra los grupos sociales objetivo. En este sentido, este tipo de campañas de desinformación suelen extenderse en el tiempo, generando un flujo continuo de eventos y narrativas que deben ser analizadas en su conjunto para poder evaluar el riesgo de radicalización. Otro factor de análisis importante es tratar de identificar a los actores sistemáticos que difunden las campañas de desinformación.

Además de utilizar un enfoque global con participación de toda la sociedad para abordar esta amenaza y promover una educación para el ciudadano con competencias adecuadas, es necesario valorar la realización de análisis anticipatorios para futuros escenarios o situaciones que puedan ser explotadas por este tipo de campañas y, de este modo, ser capaces de desplegar respuestas preventivas como el *prebunking* y la alerta de medios y verificadores.

Por último, es necesario avanzar en la protección de los actores de la sociedad civil que trabajan en la exposición y concienciación de esta amenaza, que en los últimos años se han visto expuesto a acciones de acoso, *lawfare* o incluso amenazas por parte de los actores estatales que difunden estas campañas.

Marco legal y normativo

Las características definitorias de la democracia española y del ordenamiento jurídico, principalmente el carácter normativo de la Constitución y determinados derechos fundamentales, condicionan las respuestas que puedan ofrecerse al discurso del odio desde el Derecho y los poderes públicos. España no es una democracia militante; eso implica que no se puede garantizar la libertad de expresión a través de la censura y, aunque no es un derecho ilimitado, tiene un carácter prioritario, incluso sobre otros derechos fundamentales. El disfrute de todos los derechos fundamentales exige regulaciones y medidas proporcionales y decisiones ponderadas. En consecuencia, ninguna autoridad pública puede erigirse en guardián de la verdad. La libertad de expresión es un pilar esencial de toda sociedad democrática y, por ello, han de extremarse las cautelas al decidir la eventual retirada de contenidos impuesta por autoridades administrativas y judiciales, siempre basándose en un marco normativo e institucional respetuoso con los valores y principios constitucionales.

En España existe una variedad normativa y de protocolos, tanto jurídicamente vinculante como de *soft law*, para combatir el discurso de odio *online*, ofreciendo así diversas posibilidades, desde la regulación de las plataformas y medios hasta la sanción administrativa y penal.

Destacan la Ley de Servicios Digitales (DSA) (esto es, el Reglamento (UE) 2022/2065, de 19 de octubre de 2022, relativo a un mercado único de servicios digitales por el que se modifica la Directiva 2000/31/CE, Reglamento de Servicios Digitales), la Decisión Marco 2008/913/JAI del Consejo, de 28 de noviembre de 2008, relativa a la lucha contra determinadas formas y manifestaciones de racismo y xenofobia mediante el Derecho penal, la Recomendación de la Comisión Europea contra el Racismo y la Intolerancia (ECRI) del Consejo de Europa nº 15 sobre discurso de odio del año 2016, la Recomendación para combatir el discurso de odio del Consejo de Europa (aprobado por el Comité de Ministros el 20 de mayo de 2022) y el Protocolo para combatir el discurso de odio ilegal en línea (firmado por el Ejecutivo, el Poder Judicial y la Fiscalía), además del resto del ordenamiento jurídico nacional aplicable como el Código Penal (CP).

La utilización de dichas herramientas varía en función de las distintas modalidades de discursos de odio: en primer lugar, el penalmente relevante; en segundo lugar, el no penalmente relevante, pero con respuesta administrativa o civil; y, en tercer lugar, los mensajes ofensivos sin respuesta jurídica, pero a la que hay que responder con una contranarrativa.

Así, en principio, mentir consciente o involuntariamente, es decir faltar a la verdad, no debería comportar responsabilidades penales. Conforme a lo anterior y teniendo en cuenta que uno de

los principios que rigen el Derecho Penal es el de intervención mínima, se debe excluir *a priori* de cualquier responsabilidad penal, sin perjuicio de otro tipo de responsabilidades, la divulgación involuntaria de contenidos falsos o tendenciosos.

La desinformación penalmente relevante debe reunir cuatro elementos: que sea carente de verdad; difundida deliberadamente (la motivación ha de ser discriminatoria); dolosa; y orientada a influir y manipular la opinión pública.

La hipotética respuesta normativa a este comportamiento no está lo suficientemente clara en nuestro Código Penal dentro del catálogo de conductas tipificadas en los diferentes apartados del art. 510 CP. Pero se reflexiona sobre la posibilidad de abrir, en algunos casos excepcionales y desde una perspectiva de política criminal, el delicado debate sobre la necesidad de perseguir penalmente aquellos comportamientos en los que con manifiesto y consciente desprecio a la verdad se difunden públicamente contenidos falsos o deliberadamente manipulados en los que su autor se ha hecho o se ha podido hacer un representación razonable de que de sus publicaciones van a generar entre la población reacciones de odio, hostilidad, violencia o discriminación, humillación, menosprecio etc. contra personas o grupos por motivos discriminatorios.

Esto surge porque en la actual regulación de nuestro Código Penal existen ciertas dificultades para encajar la conducta de difundir bulos o *fake news* en los tipos delictivos existentes del art. 510 del CP.

Del análisis de las conductas incardinadas en el art. 510.1.A) y el art. 510.2.A) del CP, se desprende que la difusión de bulos podría encajar mejor en los actos de lesión a la dignidad de los colectivos hacia el que se dirige el bulo del segundo de los preceptos, pero la falta de previsión específica en este artículo para estos comportamientos hace difícil su encaje, dejando un amplio margen a su interpretación.

No hay jurisprudencia sobre desinformación en el marco de delitos de odio. Los únicos dos casos juzgados hasta el momento han sido resueltos mediante sentencias de conformidad.

Para el caso de que los hechos no sean considerados delictivos, podrían constituir una infracción administrativa sancionable, lo que requiere la tramitación de un rápido expediente administrativo con el objetivo de no dejar impune estas conductas. La eficacia en la aplicación del régimen jurídico del Derecho administrativo sancionador (Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación y Ley 4/2023, de 28 de febrero, para la igualdad real y efectiva de las personas trans y para la garantía de los derechos de las personas LGTBI), exige establecer de forma inminente mecanismos ágiles de coordinación dado que la competencia sancionatoria está repartida entre las diferentes Comunidades Autónomas, a fin de garantizar una respuesta administrativa uniforme en todo el Estado, rápida y ágil; también es necesario implementar nuevas soluciones legislativas para poder perseguir los hechos cometidos por internet y redes sociales en los que en muchas ocasiones no puede establecerse un lugar concreto de comisión; por tanto, todo ello exige una autoridad administrativa sancionadora no vinculada a una sede territorial autonómica.

En muchas ocasiones para la determinación de los responsables de estas conductas se precisa la identificación del titular de la cuenta de la red social en la que se están llevando a efecto las mismas. Tal posibilidad no se ve limitada a la investigación de delitos graves pues, al tratarse de datos de abonado que no afectan al derecho fundamental al secreto de las comunicaciones y obtenidos con la finalidad de facilitar la operatividad y agilidad de las investigaciones, el art. 588 ter m) de

la LECrim permite que puedan ser directamente recabados por el Ministerio Fiscal o la Policía Judicial en cualquier clase de investigaciones y sin necesidad de autorización judicial. Quedan con ello superados los obstáculos que planteaban los operadores de comunicaciones para la entrega de esta clase de datos con el argumento de que de que los datos eran conservados junto con los datos de tráfico que están obligados a conservar en aplicación de la Ley 25/2007, de Conservación de Datos relativos a las comunicaciones electrónicas y a las redes públicas de comunicaciones - dictada en desarrollo de la Directiva 2006/24/CE posteriormente anulada por el TJUE en Sentencia a 8 de abril de 2014 – en la que se limita el acceso a los datos (de tráfico) conservados por los operadores de comunicaciones a la investigación, detección y enjuiciamiento de un delito grave exigiendo autorización judicial para su obtención.

Por su parte y en lo que se refiere al ámbito administrativo no han de olvidarse las posibilidades que ofrece el art. 10 de la DSA en base al cual, las autoridades judiciales o administrativas nacionales pertinentes, sobre la base del Derecho de la Unión aplicable o del Derecho nacional aplicable en cumplimiento del Derecho de la Unión, podrán emitir órdenes de información a los prestadores de servicios intermediarios para que informen sobre el/los destinatarios de un servicio en el marco del cual se ha publicado el discurso de odio. Dada su relevancia en esta materia el principal reto a afrontar es a agilización e impulso del desarrollo normativo interno que permita su aplicación. También ha de tenerse en cuenta las facultades que ostenta la AEPD en el marco de las investigaciones que le competen (art. 51 de la LOPD).

Cuestión distinta es la referida a la identificación de los responsables de las concretas publicaciones constitutivas de discurso de odio que hayan podido llevarse a cabo en una determinada red social en aplicación de los procedimientos sancionadores existentes en materia de discurso de odio (Ley 15/2022 o Ley 4/2023, entre otros) que sigue presentando algunos retos. Al respecto, ha de tenerse en cuenta que la obtención de las direcciones IP y otros datos de conexión vinculados a dicha publicación, al tratarse de datos de tráfico, exigen la necesidad de autorización judicial, de conformidad con lo establecido en el art. 588 ter j) de la LECrim que, por aplicación de los criterios rectores del art. 588 bis a) de la misma norma, se verá restringida a los supuestos en que dicha medida se estime proporcionada.

La relevancia de la Ley de Servicios Digitales (DSA), que pone el foco en las grandes plataformas (*Very Large online Platforms*, VLOP), está fuera de toda duda, pero para que pueda desplegar su eficacia en España es necesario que se dote de competencias a nivel estatal a la Comisión Nacional del Mercado de la Competencia (CNMC). A partir de la lectura de dicha ley, puede destacarse que el concepto de desinformación tiene tres elementos principales: en primer lugar, la falsedad del contenido; en segundo lugar, la intención de hacer daño; y, en tercer lugar, la sistemática en la propagación.

Partiendo de una base de corregulación de códigos de conducta obligatorios para las VLOP, fruto del acuerdo entre la Comisión Europea y las plataformas, el papel de la CNMC utilizando la Ley de Servicios Digitales puede contribuir de las siguientes maneras: en primer lugar, designación a los denominados alertadores fiables para la notificación de contenidos como actores privilegiados, defendiendo la libertad de expresión de los medios fiables (conectando esto con el Reglamento de Medios); en segundo lugar, dado que los contenidos son difundidos por plataformas, puede obligar a una plataforma a analizar los riesgos sistémicos, a crear fórmulas de mitigación de riesgos (con la supervisión de las agencias estatales y la agencia europea), a promocionar una acción conjunta de autocontrol y supervisión, a desmonetizar este tipo de contenidos, e incluso creando protocolos de crisis; en tercer lugar, tomando medidas sobre la transparencia en las reglas de

moderación de contenidos, ya sea prohibiendo patrones oscuros, la elaboración de perfiles sobre la base de datos protegidos para publicidad, u obligando a crear repositorios de anuncios.

El Protocolo para combatir el discurso de odio ilegal en línea es una herramienta de utilidad en lucha contra el discurso de odio en internet, del que podría destacarse la acreditación y la formación de los comunicantes fiables por parte de la Administración y las empresas proveedoras de servicios de alojamiento de datos (apartado IV), para identificar así discursos de odio (a diferencia con la DSA, por la que se puede formar, pero no acreditar). La entrada en vigor de la DSA y la creación de la nueva figura del Coordinador de Servicios Digitales exigirá la modificación del contenido del citado Protocolo para dar una nueva respuesta coordinada en la retirada de contenidos ilícitos.

De la aplicación de este protocolo puede deducirse que la mayoría del discurso de odio detectado reúne cuatro elementos: en primer lugar, el grupo de los más afectados, que en la actualidad son personas del norte de África, musulmanes, raza negra; en segundo lugar, el tipo de discurso, destacando un 53% deshumanización grave y discurso agresivo explícito un 54%; en tercer lugar, vinculación a eventos sobre inseguridad ciudadana; y en cuarto y último lugar, uso de información falsa.

Ámbito tecnológico

La tecnología juega un papel dual en el contexto de la desinformación y el discurso de odio. Por un lado, se ha constatado que las herramientas tecnológicas pueden amplificar estos problemas. Las redes sociales y las plataformas digitales son usadas para diseminar desinformación de manera rápida y masiva, fomentando la polarización social y el discurso de odio. La reciente democratización de los modelos de lenguaje avanzados, como los modelos de IA generativa, puede empeorar esta situación al facilitar la creación de contenido manipulado y falso de manera más sutil y efectiva.

Por otro lado, tenemos la responsabilidad de aprovechar estas mismas tecnologías para contrarrestar estos efectos negativos. Los avances en la inteligencia artificial y las tecnologías de procesamiento de lenguaje natural pueden utilizarse ya no solo para detectar patrones de comportamiento inapropiados (como discursos de odio, ofensividad, desinformación) sino también para mitigarlos mediante la generación de contranarrativas que nos permitan contrarrestar estos discursos de odio mediante discursos positivos que favorezcan una interacción digital mucho más respetuosa y constructiva.

La rápida evolución en los últimos años de la inteligencia artificial generativa y, en especial de los modelos lingüísticos de gran tamaño (LLM), ha centrado el foco en muchos investigadores que trabajan en la detección de discurso en línea y en la creación de contranarrativas.

Sin embargo, el uso de estos modelos para la detección del discurso de odio plantea algunos retos, como la necesidad de contar con modelos entrenados en los idiomas de interés (castellano y lenguas cooficiales), de avanzar en la detección de sarcasmo e ironía, y en la adaptabilidad a la ambigüedad del lenguaje y la evolución temporal del mismo, así como a la diversidad lingüística regional. Por otro lado, existen algunos factores que se consolidan como oportunidades para el uso de estos modelos, entre ellos la existencia de numerosos recursos en español para poder entrenar estos modelos, la capacidad de los mismo de tener un aprendizaje continuo, o la cada vez mayor cooperación multidisciplinar de expertos, desde sociólogos y psicólogos hasta especialistas en el ámbito del Derecho.

Los elevados niveles de información que se observan actualmente en las redes sociales hacen que la detección del discurso de odio automatizada no sea suficiente ya que, aunque se detecte, no existen suficientes recursos para hacerle frente de forma manual. Por ello, estos modelos de inteligencia artificial también están siendo probados para las labores de mitigación y generación automática de contranarrativas. De esta forma, se elabora una alternativa a la del bloqueo o eliminación de estos mensajes que es menos lesiva con la libertad de expresión y se integra como un posible mecanismo en la prevención de la radicalización.

La generación de contranarrativas se basa en la creación automática o supervisada de respuestas a mensajes de odio en las que se ofrece una visión alternativa y positiva. Actualmente se han estudiado diferentes estrategias de contranarrativas, si bien no todas tienen la misma efectividad, dependiendo de cada situación y del público al que van dirigidas. Del mismo modo que los actores hostiles que promueven o difunden campañas de desinformación explotan las vulnerabilidades existentes en las sociedades objetivo, la defensa frente a la desinformación y la influencia hostil en el entorno de información mediante narrativas positivas debe partir de un análisis, no sólo de nuestras vulnerabilidades y de qué nos divide, sino también de nuestras fortalezas y de aquello que nos mantiene unidos y cohesionados como sociedad democrática, plural y respetuosa con las diferencias en las características identitarias existentes.

Actualmente, estos diseños requieren de una implementación más extensa y de la necesidad de mejorar los mecanismos para evaluar la efectividad de las contranarrativas.

Por otro lado, aunque poco a poco se ha avanzado en el estudio de los algoritmos de promoción y recomendación de contenido de las plataformas, es necesario seguir progresando para entender las implicaciones para el fomento de la polarización, así como comprender cómo las dinámicas de sobreestimulación que promueven los modelos actuales de atención de algunas redes impactan negativamente en los mecanismos cognitivos necesarios para la correcta evaluación e integración de la información.

Alfabetización mediática, concienciación y el papel del tercer sector

La alfabetización mediática e informacional es un reto esencial para combatir eficazmente la desinformación. En el ámbito académico, no es solo importante incluir en el BOE un *currículum* sobre desinformación sino promover los esfuerzos para que llegue de forma efectiva a las aulas.

Además, la capacidad actual de la alfabetización en el ámbito educativo no es suficiente, dado que la capacidad de actuar de la academia es limitada. En el actual entorno informativo los medios de comunicación y las plataformas digitales son los que tienen mayor capacidad de llegar a todos los públicos.

La formación a la ciudadanía, y en especial a los jóvenes, ha de incidir en mejorar su capacidad para detectar bulos, trabajando el pensamiento crítico, el análisis del contexto, la evaluación con indicadores, previamente facilitados, para que sirvan para analizar los condicionamientos a la credibilidad de la fuente. Esto aplica tanto al medio escrito como al audiovisual, e incluso al acústico.

También el avance en el conocimiento de las diferentes tácticas utilizadas en la difusión de discurso de odio y desinformación son esenciales para informar la alfabetización mediática.

Así, por ejemplo, los mensajes que contienen un componente de odio son más propensos a ser compartidos; sin embargo, tienden a disminuir su credibilidad en general. Este factor es relevante a la hora de avanzar en el plano de la concientización y alfabetización.

Por otro lado, el rol a desempeñar por las familias, en colaboración con los docentes es esencial para facilitar el trabajo en las aulas de la aceptación del desacuerdo, el disenso, la diversidad de perspectivas, del conflicto, inherente a las relaciones sociales, para ayudar a confrontar y aun a prevenir el discurso de odio en las redes sociales. La educación se ha mostrado eficaz para interrumpir el proceso que activa el odio y la violencia.

También la verificación de la comunicación, cualquiera que sea el medio que se utilice, se ha demostrado como un elemento eficaz para detectar y neutralizar la desinformación desapercibida, aunque requiere de una alta especialización, recursos y tiempo. Además, no es suficiente con el señalamiento o etiquetado como desinformación, sino que es necesario explicar por qué es desinformación. La colaboración entre verificadores y la transparencia metodológica a la hora de trabajar estos asuntos también son clave. En este sentido, mecanismos como las notas de comunidad o sistemas de etiquetado de contenido fruto de trabajo colectivo de los usuarios agregando contexto o verificaciones en contenido engañoso, es un complemento de la moderación de contenido que debiera potenciarse.

A su vez, los profesionales, no solo de ciencias de la información, sino de todo tipo de medios de comunicación social, salvaguardada la libertad de expresión, precisan de ejercicios de formación y sensibilización sostenidos en el tiempo, con enfoque de interculturalidad e interseccionalidad. Es importante incluir a los educadores en los debates en la formulación de estrategias, no solo de alfabetización mediática sino, también, de comunicación. Con esto se puede tener en cuenta y valorar también los efectos a largo plazo de las mismas. También es necesario concienciación en el ámbito empresarial para limitar la monetización de este tipo de campañas, eliminando o limitando la motivación lucrativa.

Actualmente, las organizaciones del tercer sector carecen de recursos y procedimientos comunes para abordar esta amenaza. La cooperación con otros actores de la sociedad civil, como los verificadores, ha dado buenos resultados para entender mejor esta amenaza.

En este sentido, la DSA constituye el instrumento preciso en la actualidad para articular la colaboración responsable triangulada, precisa entre las autoridades nacionales, los prestadores de servicios digitales globales, marcadamente, las redes sociales pero no exclusivamente, y las organizaciones de la sociedad civil, idóneamente a través de plataformas que las aglutinen y las empoderen, para trascender a los procesos participativos actuales, y posicionarse como prescriptores estratégicos a considerar, para colaborar en establecer las bases de las políticas públicas y argumentarios especializados para prevenir y luchar contra la desinformación y el discurso de odio.

Por último, y debido a la dimensión internacional de la desinformación, es importante colaborar con el resto de países para que impulsen mecanismos de concientización y alfabetización, así como promover que las herramientas de detección y los recursos de moderación en los diferentes idiomas sean suficientes. Con ello, se avanzará en limitar el impacto que estas campañas pueden tener sobre los colectivos en terceros países y también limitar su difusión, evitando que se trasladen a nuestro país por la permeabilidad de las redes o, incluso, a través de la diáspora. A la hora de impulsar acciones de comunicación o alfabetización en otras regiones es esencial atender a las circunstancias culturales y sociales de aquellas, a fin de evitar amplificar el propio efecto de la desinformación.

RECOMENDACIONES

Es necesario avanzar en el plano de la **concienciación** para entender y documentar mejor esta amenaza y sus efectos y para tener una base para implicar a todos los actores de la sociedad que deben aunar esfuerzos para luchar contra la misma.

Las narrativas están identificadas. Las estrategias también comienzan a estar identificadas. La identificación plena de ambas (narrativa más estrategia) permitirá identificar de manera cada vez más elaborada la intencionalidad *prima facie* (especialmente si usamos inteligencia artificial para ello). Esto puede suponer un avance muy significativo para la lucha contra los discursos de odio porque tendremos uno de los elementos más difíciles (la intención del autor) en su definición. Obviamente en caso de controversia será solo una identificación provisional susceptible de revisión judicial posterior.

Parece prometedor proceder a la identificación de los "actores". Tanto más posible si hemos desarrollado el punto anterior (identificación de narrativas, estrategias e intencionalidad). Podría explorarse la posibilidad de que los actores habituales de desinformación para creación de discurso de odio puedan ser tratados como algo parecido a "organización criminal".

En el **ámbito normativo**, será necesario valorar una reforma del art. 510 CP para incluir la persecución penal de la difusión pública y maliciosa de contenidos manifiestamente falsos en las condiciones expuestas anteriormente. En este sentido, debería valorarse la posibilidad de abrir el debate jurídico, sin prejuzgar un resultado concreto, tanto a nivel nacional como europeo, sobre la oportunidad desde una perspectiva de política criminal de crear nuevo tipo delictivo específico, con todas las cautelas y limitaciones apuntadas, para perseguir penalmente aquellos comportamientos en los que, de manera concertada y coordinada, con manifiesto y consciente desprecio a la verdad se difunden públicamente contenidos falsos o deliberadamente manipulados en los que su autor se representa o pueda representarse razonablemente que de sus publicaciones se van a generar entre la población reacciones de odio, hostilidad, violencia o discriminación, humillación o menosprecio contra personas o grupos por motivos discriminatorios, todo ello con el fin de evitar una interpretación forzada del actual delito de lesión de la dignidad por motivos discriminatorios previsto en el art. 510.2.A) del CP, dado que existen verdaderas dificultades para incluir la conducta consistente en difusión de bulos. La redacción del nuevo delito debería limitarse a los supuestos de intencionalidad directa de la acción de expandir bulos o de llevar a cabo la acción con dolo eventual o culpa consciente para no dejar atípica ninguna conducta.

No obstante, no se deben usar herramientas excesivas en la respuesta sancionatoria ni en la posible amenaza a los propios valores defendidos. Una respuesta colaborativa y graduada parece mucho más aconsejable. Se debería poder acudir a la vía administrativa para los hechos que no sean graves y que no supongan una puesta en peligro del colectivo al que atacan.

Es necesario mejorar en la regulación y aplicación del régimen administrativo sancionador y en la posibilidad de cesión de datos de tráfico para procedimientos administrativos sancionadores y civiles de vulneración del derecho al honor, intimidad o propia imagen, así como en la coordinación entre instituciones estatales y autonómicas para garantizar la efectiva imposición de sanciones y evitar su prescripción. Se han de implementar soluciones para garantizar una respuesta coordinada, uniforme y útil en el ámbito administrativo sancionador que implique a todas las administraciones estatales y autonómicas para limitar el alcance y efectos de estas amenazas; en

todo este proceso, penal y administrativo, deberá jugar un papel clave el Coordinador de Servicios Digitales en el marco del Reglamento de Servicios Digitales en la retirada de contenidos ilícitos penales y administrativos para lo cual se tienen que efectuar las habilitaciones normativas precisas de forma urgente.

Por último, es necesario implementar de forma urgente las previsiones de la DSA (Reglamento 2022/2065 del Parlamento Europeo y del Consejo relativo a un mercado único de servicios digitales) y del protocolo de retirada de contenidos ilícitos en internet.

Por otro lado, uno de los principales retos **en el campo tecnológico** es la transferencia de las diferentes investigaciones que se están desarrollando en el marco de la detección y respuesta a estas amenazas. Es necesario identificar posibles oportunidades para la implementación efectiva de estas tecnologías.

La utilización de herramientas de inteligencia artificial en la lucha contra la desinformación vinculada al discurso de odio parece la vía más prometedora, pero todavía no está suficientemente desarrollada. Se plantean dificultades técnicas y jurídicas. De ahí se deducen dos necesidades: en primer lugar, incrementar los esfuerzos en términos de recursos materiales y humanos para el diseño de herramientas eficaces; en segundo lugar, asegurar la interdisciplinariedad en el diseño para la inclusión de los valores éticos y jurídicos que se pretenden defender.

El uso de herramientas de inteligencia artificial requiere, a su vez, una supervisión adecuada de éstas y su consideración como herramientas de alto riesgo. Eso supondría cubrir exigencias técnicas (Reglamento de IA de la UE) y establecer los procedimientos para que los poderes públicos las autoricen, reconozcan, implanten y supervisen (gobernanza de la lucha con inteligencia artificial contra los discursos de odio basados en desinformación).

En el ámbito de la **alfabetización mediática**, es esencial identificar estrategias para garantizar que esta llegue realmente a los ciudadanos. Para ello es necesario implicar a los medios de comunicación y a las plataformas digitales, así como garantizar que los currículos académicos sobre desinformación se implanten de forma efectiva en todas las aulas.

Además, sería interesante no sólo una contranarrativa sino también un enfoque de exposición pública o contramarca, revelando a la sociedad las operaciones de influencia, desinformación e intentos de manipulación a la población española.

Por último, es necesario avanzar en la formación de las **organizaciones del tercer sector**, explorando mecanismos de cooperación con otros actores como los verificadores, la academia o las plataformas digitales. Además, es interesante fomentar el desarrollo de procedimientos comunes que faciliten el trabajo de las organizaciones en esta materia.



CONFERENCIA CAMPAÑAS DE DESINFORMACIÓN Y PROMOCIÓN DEL DISCURSO DEL ODIO

Aula Polivalente II de la Facultad de Derecho de la Universidad Complutense de Madrid
Pl. Menéndez Pelayo, 4, Madrid

MIÉRCOLES 18 DE SEPTIEMBRE DE 2024

- 09:45 **Bienvenida**
- 10:00 **Definición y valoración de la amenaza**
- Beatriz Marín, Servicio Europeo de Acción Exterior.
- 11:45
- Ruben Arcos, Universidad Rey Juan Carlos de Madrid.
 - Alicia Moreno Delgado, Universidad Internacional de La Rioja.
 - Raquel Godos, EFE Verifica.
 - Alejandro González, Departamento de Seguridad Nacional.
- 12:00 **Marco legal y normativo**
- Miguel Ángel Aguilar, Fiscal de Sala Coordinador de la Unidad de los Delitos de Odio y Discriminación de la Fiscalía General del Estado.
- 13:45
- Julio del Valle de Iscar, Director General para la Igualdad Real y Efectiva de las Personas LGTBI+ del Ministerio de Igualdad.
 - Karoline Fernández de la Hoz Zeitler, Directora del Observatorio Español del Racismo y la Xenofobia del Ministerio de Inclusión, Seguridad Social y Migraciones.
 - Carlos Aguilar Paredes, Comisión Nacional de los Mercados y la Competencia.
 - Alfonso Peralta Gutiérrez, Juez de Primera Instancia e Instrucción.
 - Rafael Bustos Gisbert, Catedrático de Derecho Constitucional de la Universidad Complutense de Madrid.
- Pausa comida*
- 15:45 **Retos y oportunidades de las nuevas tecnologías**
- Emilio Delgado López Cózar, Universidad de Granada.
- 17:30
- David Blanco Herrero, University of Amsterdam (Países Bajos).
 - Gavin Abercrombie, Heriot Watt University (Reino Unido).
 - Flor Miriam Plaza del Arco, Bocconi University (Italia).
 - Maite Martín Valdivia, Universidad de Jaén.
 - Carlos Arcila Calderón, Universidad de Salamanca.
- 15:45 **Comunicación estratégica, alfabetización mediática y el papel del tercer sector**
- Manuel Gértrudix Barrio, Universidad Rey Juan Carlos de Madrid..
- 19:10
- Alberto Izquierdo Montero, UNED.
 - Pablo Hernández Escayola, Maldita.es.
 - Natalia Sancha, Servicio Europeo de Acción Exterior.
 - Marisa Gómez, Plataforma de ONG de Acción Social.
 - Carmen Girón Tomás, Doctoranda en Derecho y Ciencias Sociales, UNED.

ANEXO: Agenda de la conferencia celebrada el 18 de septiembre de 2024

REFERENCIAS BIBLIOGRÁFICAS

Alonso-Villota, M., & Arcos, R. (2024). The Coercion-Manipulation-Persuasion Framework: Analyzing the Modus Operandi of Systems of Non-State Actors. *Terrorism and Political Violence*. doi:10.1080/09546553.2024.2357082

EEAS Stratcom Division. (2023). FIMI targeting LGBTIQ+ people: Well-informed analysis to protect human rights and diversity. Servicio Europeo de Acción Exterior. Obtenido de <https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-LGBTQ-Report.pdf>

Hoogensen Gjørsv, G., & Jalonen, O. (2023). Identity as a tool for disinformation: Exploiting social divisions in modern societies. Centro Europeo de Excelencia para la Lucha contra las Amenazas Híbridadas. Obtenido de <https://www.hybridcoe.fi/wp-content/uploads/2023/11/20231108-Hybrid-CoE-SA-34-Identity-as-a-tool-for-disinformation-WEB.pdf>

Red de Concienciación sobre la Radicalización (RAN). (2020). The Impact of Conspiracy Narratives on Violent RWE and LWE Narratives. Comisión Europea. Obtenido de https://home-affairs.ec.europa.eu/system/files/2021-01/ran_c-n_concl_pap_impact_consp_narr_on_vrwe_vlwe_24-25_112021_en.pdf

Szakács, J., & Bognár, É. (2021). The impact of disinformation campaigns about migrants and minority groups in the EU. Parlamento Europeo. Obtenido de [https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/653641/EXPO_IDA\(2021\)653641_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/653641/EXPO_IDA(2021)653641_EN.pdf)

