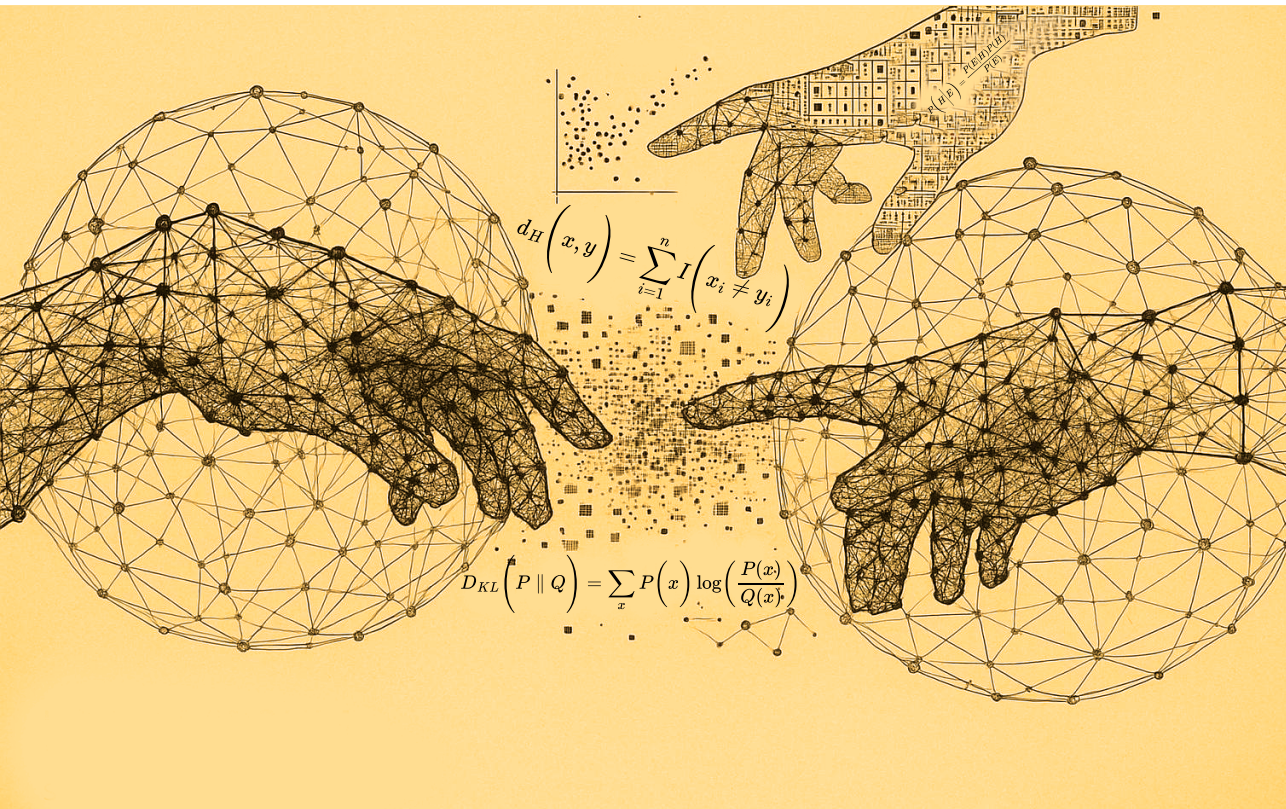
FORO CONTRA LAS CAMPAÑAS DE DESINFORMACIÓN

INICIATIVAS 2025: CONCLUSIONES Y RECOMENDACIONES DE LOS EXPERTOS

ESTRATEGIAS PARA EL SEGUIMIENTO DE ACTORES: MÉTODOS Y ANÁLISIS TÉCNICOS EN LAS OPERACIONES DE DESINFORMACIÓN



ESTRATEGIAS PARA EL SEGUIMIENTO DE ACTORES: MÉTODOS Y ANÁLISIS TÉCNICOS EN LAS OPERACIONES DE DESINFORMACIÓN

Todos los expertos participantes en los Grupos de Trabajo, tanto del sector público como del privado, 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 ni de las organizaciones o entidades públicas y privadas representadas, quienes no necesariamente comparten todas las conclusiones o propuestas.

ÍNDICE

ESTRATEGIAS PARA EL SEGUIMIENTO DE ACTORES: MÉTODOS Y ANÁLISIS TÉCNICOS EN LAS OPERACIONES DE DESINFORMACIÓN

COORDINADORES

Coordinadores:

Iván Portillo Morales

Expertos del Ministerio del Interior – Oficina de Coordinación de Ciberseguridad (OCC)

Autores y colaboradores:

Alejandro Bruquetas Gómez

Ángel Luis Veloy Mora

Claudia Sánchez-Girón López

David Arroyo Gaurdeño

Gonzalo Terciado Terciado

José Manuel Ávalos Morer

Pablo Bentanachs González

Paula González Nagore

Pedro Gómez García

Roberto Lara González

Expertos del Cuerpo Nacional de Policía (Comisaría General de Información)

Expertos de la Guardia Civil (Jefatura de Información)

METODOLOGÍA TÉCNICA APLICADA A LA DESINFORMACIÓN

La desinformación se ha consolidado como un fenómeno complejo que amenaza seriamente el ecosistema digital actual. Su capacidad para incidir en la opinión pública, alterar procesos políticos, desestabilizar economías y erosionar la confianza en las instituciones democráticas la convierte en un objeto de estudio prioritario desde perspectivas tanto técnicas como estratégicas. Su naturaleza adaptativa, potenciada por las nuevas tecnologías como la utilización de inteligencia artificial generativa, la automatización de la difusión y la explotación de vulnerabilidades en las redes sociales configuran un escenario donde las respuestas tradicionales resultan claramente insuficientes, lo que provoca una mayor dificultad tanto para su detección como su neutralización.

Ante este escenario, se requiere una metodología que no solo identifique narrativas manipuladas, sino que también mapee canales de difusión, examine mecanismos de sostenimiento económico, rastree redes de influencia y analice el modus operandi de los actores. El propósito es transformar información fragmentada en inteligencia accionable, que permita diseñar estrategias de mitigación frente a las tácticas, técnicas y procedimientos empleadas, además de establecer un estándar común que facilite la cooperación interinstitucional.

Planificación e identificación de fuentes

La fase de planificación es fundamental, pues define qué observar, en qué entornos digitales y con qué propósito, evitando la recolección indiscriminada de datos. Su objetivo es vincular motivaciones, actores y escenarios de actuación, con especial atención a fuentes hispanohablantes y, en particular, a aquellas con foco en España, ya que en este ámbito confluyen narrativas globales con discursos adaptados específicamente a audiencias locales.

El mapa de difusión abarca tanto medios estatales como aliados. Los primeros incluyen portales digitales, televisiones online y perfiles oficiales en redes sociales que funcionan

como instrumentos de proyección de narrativas vinculadas a intereses gubernamentales de países extranjeros. Los segundos, aunque no dependen directamente de un Estado, actúan como satélites narrativos que refuerzan y diversifican las fuentes de propaganda.

Las redes sociales, como *X*, *Facebook* e *Instagram*, ocupan un papel central en la viralización de contenidos. No solo funcionan como espacios de consumo informativo, sino que también se consolidan como plataformas clave para la propagación de bulos y narrativas manipuladas, especialmente aquellas vinculadas al ámbito político y social.

En paralelo, las plataformas de *streaming y vídeo*, entre ellas *YouTube*, *Twitch*, *DLive*, *BitChute* y *Odysee*, permiten desplegar estrategias audiovisuales de gran alcance. YouTube, por ejemplo, combina la publicación de documentales manipulados con transmisiones políticas en directo; Twitch conecta con audiencias jóvenes a través de contenidos interactivos; mientras que BitChute y Odysee funcionan como refugios para materiales excluidos de servicios más regulados. A ello se suman las aplicaciones de mensajería instantánea, como *Telegram* y *Discord*, que operan como nodos críticos de coordinación y distribución, ofreciendo entornos donde convergen la inmediatez en la difusión y la protección del anonimato mediante identidades digitales alternativas.

Por su parte, las comunidades digitales y los foros especializados, como Reddit o espacios temáticos, contribuyen a la formación de grupos cohesionados en torno a narrativas concretas. No obstante, los denominados *foros underground* operan como espacios semiclandestinos orientados al intercambio de recursos técnicos, herramientas y estrategias orientadas a la manipulación.

Por último, los sitios web propios de los actores cumplen un rol estratégico, al servir como repositorios de contenidos y *hubs* narrativos desde los cuales se controlan los marcos discursivos y se articulan enlaces hacia otros espacios.

Un aspecto clave en este entramado es la identificación de los mecanismos de financiación que sostienen estas operaciones. Entre ellos se incluyen el *micromecenazgo*, a través de plataformas como *Patreon*, *BuyMeACoffee*, *Ko-fi*, la monetización nativa en redes sociales como YouTube, Twitch, X, así como las tiendas de *merchandising*, las donaciones directas en portales web (por medio de *PayPal*, *GabPay* u otros servicios similares) y el uso de criptomonedas como *Bitcoin*, *Ethereum* o *Monero* que ofrecen un mayor grado de anonimato y una trazabilidad limitada. Aunque con menor frecuencia, en algunos casos se detectan campañas de *crowdfunding* específicas, ya sea en plataformas dedicadas como *Kickstarter* o a través de los propios sitios web de los actores.

El cruce entre los canales de difusión y los modelos de financiación muestra que la desinformación no responde únicamente a motivaciones ideológicas o políticas, sino que

también se sostiene en incentivos económicos que refuerzan su persistencia y capacidad de expansión.

Obtención y análisis de la información

La fase de obtención y análisis de la información tiene como objetivo transformar la actividad que se desarrolla en los distintos entornos digitales en datos estructurados y útiles para la investigación. Para ello, la información recopilada se somete a un proceso de normalización y homogeneización dentro de un marco temporal, se eliminan duplicados y se enriquece con información complementaria. Posteriormente, se clasifica según su formato y su nivel de verificación, asegurando así que el análisis se realice sobre bases sólidas y comparables.

En este marco, el análisis textual constituye una parte central del proceso. A través del estudio de palabras clave, hashtags y patrones de uso del lenguaje, se identifican los principales temas de conversación y los marcos retóricos que los sustentan, como los discursos de amenaza, conspiración o declive. Asimismo, se evalúa la orientación de los mensajes frente a determinados temas, distinguiendo entre posiciones favorables, contrarias o neutrales, lo que permite mapear el posicionamiento discursivo de los actores implicados.

Así mismo, el estudio de las redes sociales ofrece una perspectiva relacional que permite reconocer comunidades, identificar líderes de difusión y detectar posibles dinámicas de coordinación. Mediante el análisis de las conexiones entre usuarios, contenidos y etiquetas, se revelan comunidades cohesionadas, nodos puente entre plataformas y comportamientos que pueden sugerir la existencia de automatizaciones o estrategias coordinadas de amplificación.

Por otro lado, el trabajo con imágenes, vídeos y audios introduce una capa analítica adicional. En el caso de las imágenes se utilizan búsquedas inversas y revisiones de metadatos para determinar su origen, variantes y posibles manipulaciones. En cuanto a los vídeos y audios se examinan fragmentos clave, subtítulos y descripciones con el objetivo de identificar ediciones o alteraciones que modifican el sentido original del contenido.

A este conjunto de procedimientos se suma el análisis financiero, orientado a comprender las estructuras que sostienen estas campañas. El rastreo de criptomonedas, plataformas de micromecenazgo y sistemas de donaciones online permite descubrir vínculos económicos entre actores y relacionar picos de ingresos con momentos de gran actividad narrativa. Esta correlación entre los flujos económicos y la intensidad

comunicativa puede ofrecer indicios relevantes para la atribución de responsabilidades y la identificación de redes de apoyo material.

En última instancia, el resultado de esta fase se traduce en la elaboración de productos analíticos como informes, mapas de redes y fichas narrativas, que sirven de base para identificar las tácticas, técnicas y procedimientos (TTP) empleados en las operaciones de desinformación.

Estrategias de identificación y mitigación de TTP

La identificación de TTP constituye un paso fundamental en la metodología, ya que permite traducir los hallazgos obtenidos en fases previas en conocimiento estructurado y replicable. El uso de marcos como DISARM ofrece un esquema metodológico que facilita desagregar las campañas de desinformación en componentes específicos, transformando un fenómeno complejo en procesos observables, comparables y predecibles.

Más que enumerar técnicas, el análisis de TTP busca detectar patrones y dinámicas recurrentes. La combinación reiterada de *bots* con la amplificación en mensajería instantánea, o la correlación entre picos de financiación en criptomonedas y la intensificación de narrativas audiovisuales, son ejemplos que permiten reforzar la atribución y anticipar escenarios futuros.

La metodología técnica aplicada al estudio de la desinformación se configura como un marco que articula la planificación estratégica, el análisis técnico, la trazabilidad financiera, la identificación de TTP y el diseño de estrategias de mitigación. Su principal valor reside en la capacidad de transformar los hallazgos en inteligencia accionable, orientada a diseñar un plan de respuesta eficaz frente a un ecosistema digital caracterizado por su adaptabilidad, complejidad y persistencia.

IDENTIFICACIÓN Y CLASIFICACIÓN DE ACTORES EN OPERACIONES DE DESINFORMACIÓN

El seguimiento de actores asociados a ciberamenazas resulta cada vez más relevante, dado que varios de ellos han ampliado progresivamente su modus operandi hacia actividades de desinformación. Entre los más destacados se encuentran las Amenazas Persistentes Avanzadas (APT), habitualmente vinculadas a intereses estatales y capaces de articular operaciones que combinan ciberespionaje con campañas de influencia; los grupos hacktivistas, que complementan su acción ideológica o reivindicativa con estrategias de manipulación informativa; y los actores de cibercrimen, que recurren a la desinformación como recurso para obtener beneficios económicos o consolidar su posicionamiento dentro del ecosistema criminal.

El análisis de estas dinámicas permite comprender la convergencia entre el ámbito de la ciberseguridad y el de la desinformación, al tiempo que refuerza la atribución, la anticipación frente a las amenazas y el diseño de respuestas adaptadas a la naturaleza cambiante de cada actor.

Amenazas persistentes avanzadas (APT)

Las operaciones de desinformación impulsadas por actores estatales se han convertido en instrumentos estratégicos dentro de la guerra híbrida, empleados para moldear percepciones, debilitar consensos políticos y socavar la cohesión social. En este ecosistema, los grupos APT actúan como brazos operativos de los intereses geopolíticos de los Estados, integrando la dimensión técnica y la narrativa. Las actividades asociadas a las APT suelen estar impulsadas por actores categorizados como estado-nación o contar con su patrocinio directo, lo que les confiere una orientación clara hacia objetivos geoestratégicos. Su eficacia radica en la complejidad de sus operaciones apoyándose de la integración de acciones técnicas con operaciones de influencia. De este modo, los APT no solo buscan comprometer sistemas o exfiltrar información, sino también generar inteligencia estratégica y construir narrativas por medio de filtraciones y manipulaciones utilizadas en campañas de desinformación a gran escala.

Entre los principales actores destacan Rusia y China, con estrategias consolidadas de influencia global, mientras que Irán ha emergido con relevancia en el contexto europeo, especialmente por su capacidad de influir en países hispanohablantes como España.

Dinámicas estatales de desinformación: Rusia

Los grupos vinculados a Rusia se destacan por su elevada actividad en la ejecución de operaciones hibridas, caracterizadas por la combinación de ciberataques, filtraciones de información y manipulación narrativa. En este contexto, estructuras como la *Social Design Agency* (SDA) operan como núcleos de coordinación para campañas de desinformación, entre las que destaca "*Operation Undercut*", activa desde finales de 2023, que utiliza más de 500 cuentas falsas junto con vídeos e imágenes generados por inteligencia artificial. El propósito fundamental de estas operaciones es desacreditar a Ucrania, erosionar el consenso europeo y explotar divisiones políticas internas tanto en Estados Unidos como en la Unión Europea. Asimismo, la estrategia de injerencia rusa ha orientado parte de sus esfuerzos hacia eventos internacionales de gran repercusión, como los Juegos Olímpicos de París 2024. Este tipo de acontecimientos ofrece una plataforma ideal para amplificar el alcance de sus narrativas y captar la atención de una audiencia global, aprovechando el aumento de cobertura mediática que generan (Insikt Group, 2024a).

Las campañas de desinformación vinculadas a intereses rusos no se centran exclusivamente en España, pero sus efectos se proyectan sobre el país mediante la difusión de narrativas ampliamente empleadas en el espacio europeo. Entre ellas destacan los mensajes que subrayan los efectos negativos de las sanciones económicas sobre los Estados miembros y buscan fomentar la división social y política en el seno de la Unión Europea (EU vs Disinfo, 2023).

El contexto informativo español, caracterizado por un deterioro significativo de la confianza en los medios y por un aumento sostenido de la polarización, constituye un entorno especialmente vulnerable a este tipo de dinámicas (Badillo, 2024). En este escenario, los contenidos desinformativos encuentran condiciones favorables para amplificar tensiones y erosionar la credibilidad institucional. Esta vulnerabilidad se ha puesto de manifiesto en episodios en los que actores alineados con intereses rusos han tratado de aprovechar situaciones de especial impacto interno para promover la desconfianza hacia las instituciones públicas, proyectar una imagen de inestabilidad y cuestionar el apoyo a Ucrania (González, 2025). Además, diversos análisis sitúan a España dentro de un marco europeo donde se observan casos de desinformación relacionados con la guerra de Ucrania, acompañados de un clima de elevada desconfianza hacia el sistema mediático y de discursos crecientemente polarizados (Moreno-Castro et al., 2023).

Dinámicas estatales de desinformación: China

Las operaciones de información atribuidas a actores alineados con China han ampliado su alcance en los últimos años, tal como muestra la actividad de la red "*Empire Dragon*", identificada como una operación coordinada y probablemente vinculada a intereses gubernamentales chinos. Activa desde comienzos de 2021, esta red desarrolla campañas destinadas a influir en audiencias globales mediante la difusión de contenidos en múltiples idiomas, plataformas y temáticas (Insikt Group, 2023). A lo largo de su evolución, *Empire Dragon* ha adaptado su enfoque y, tras agosto de 2022, amplió su atención hacia Estados Unidos y sus aliados en respuesta a acontecimientos geopolíticos y al surgimiento de narrativas conspirativas.

En este contexto, la red ha difundido teorías conspirativas relacionadas con el sabotaje del gasoducto *Nord Stream* y con la supuesta existencia de programas estadounidenses de armas biológicas en Ucrania, incorporándolas entre los contenidos que circulan a través de sus canales. Del mismo modo, se observa una creciente convergencia entre las narrativas empleadas por *Empire Dragon* y aquellas promovidas por campañas de desinformación de origen ruso, lo que refleja una aproximación progresiva en los marcos discursivos utilizados por ambos actores en el entorno informativo internacional (Insikt Group, 2023).

La influencia informativa de China en España se manifiesta en la presencia de contenidos en español producidos por medios estatales chinos y difundidos a través de distintos actores del sistema mediático nacional. Entre estos se encuentran publicaciones elaboradas por órganos oficiales de comunicación chinos, como *China Watch* y *China Hoy*, junto con contenidos procedentes de Xinhua, la agencia estatal china, que han sido redistribuidos por EFE, agencia nacional española, además de la disponibilidad de canales como CGTN o China FM en plataformas de radio y televisión (Freedom House, 2022).

Esta producción informativa, alineada con los intereses y enfoques del Partido Comunista Chino, actúa como un instrumento para amplificar posicionamientos de política exterior y promover marcos de interpretación compatibles con los modelos de gobernanza defendidos por Pekín, mediante una estrategia que incluye la difusión de un elevado volumen de contenidos dirigidos a audiencias hispanohablantes (Moreno, 2024).

Dicha estrategia persigue reforzar la imagen internacional de China y proyectar narrativas favorables en debates relevantes para la opinión pública española, como los relativos al comercio, la tecnología, los derechos humanos o diversas cuestiones de política exterior. Estos objetivos se enmarcan en iniciativas más amplias orientadas a mejorar la

percepción de China en Europa y a consolidar de manera progresiva su influencia en los espacios informativos del continente (Aldama, 2020; Benedicto, 2025; Freedom House, 2022).

Dinámicas estatales de desinformación: Irán

Irán mantiene una actividad consolidada en el plano de la obtención de información sensible y de la intervención en el entorno comunicativo internacional, a través de operaciones atribuidas a grupos estrechamente vinculados a estructuras estatales. Entre ellos destacan APT35 (Charming Kitten, Mint Sandstorm) y APT33 (Peach Sandstorm, Elfin), relacionados con el Cuerpo de la Guardia Revolucionaria Islámica y responsables de campañas basadas en phishing dirigido, ataques de fuerza bruta y el uso de herramientas avanzadas de intrusión, incluido el backdoor "Tickler" y diversas infraestructuras fraudulentas orientadas al compromiso de organizaciones estratégicas (Microsoft, 2023a, 2023b, 2024; Newman, 2024).

A estas capacidades técnicas se añaden operaciones de suplantación e ingeniería social dirigidas contra investigadores, académicos y entidades de alto valor, subrayando su función dentro del aparato ofensivo iraní (Gatewatcher, 2025; Lakshmanan, 2025).

Además de estas actividades, dichos actores han ampliado su intervención hacia acciones orientadas a influir en percepciones públicas y a moldear narrativas en contextos de elevada tensión geopolítica. Diversas investigaciones indican que estos grupos combinan obtención de información, alteración de sistemas y difusión estratégica de contenidos para reforzar posiciones alineadas con los intereses iraníes, especialmente en escenarios marcados por escaladas regionales como los ataques contra instalaciones iraníes en junio de 2025 (CyberProof, 2025). En conjunto, estas dinámicas reflejan un uso coordinado de técnicas de intrusión, suplantación y circulación de mensajes manipulados con el fin de proyectar influencia, debilitar la credibilidad de actores adversarios y fortalecer la posición internacional de Irán en un entorno caracterizado por una creciente competencia estratégica (Microsoft, 2023b, 2024; Gatewatcher, 2025; Lakshmanan, 2025). Paralelamente, diversos análisis subrayan que las actividades iraníes de proyección informativa en entornos digitales combinan la amplificación automatizada de contenido en redes sociales y la utilización de identidades digitales falsas para moldear opiniones políticas y fomentar la desconfianza pública (Sha, 2025; Global Influence Operations, 2025).

En el ámbito europeo, se han señalado operaciones atribuidas a servicios iraníes orientadas a intimidar, acosar o atentar contra disidentes y personas críticas con el régimen, en ocasiones mediante redes delictivas empleadas como intermediarios operativos (Lee,

2025; England, 2025). Estas acciones se complementan con campañas de presión y deslegitimación dirigidas contra periodistas y activistas, como demuestra la campaña de amenazas y hostigamiento dirigida contra la defensora de derechos humanos *Narges Mohammadi* (Reporters Without Borders, 2025). Asimismo, se han detectado intrusiones, filtraciones selectivas y amplificación destinadas a influir y obtener réditos políticos y mediáticos (Ribeiro, 2025; CyberProof, 2025).

En España, no se han identificado campañas directamente atribuibles a actores iraníes, aunque sí se han registrado indicios de actividad indirecta que encajan en una estrategia más amplia de presión extraterritorial. El principal caso es el intento de asesinato del exdiputado *Alejo Vidal-Quadras* en 2023, una línea de investigación que apunta a una posible motivación relacionada con su posicionamiento público frente al régimen iraní, si bien no existe confirmación oficial sobre la autoría estatal (The Diplomat in Spain, 2023; Reuters, 2025; Kia, 2025).

Estas dinámicas confirman la evolución de Irán como un actor híbrido de alcance global, capaz de integrar de manera coordinada la presión política, la manipulación informativa y las operaciones en el ciberespacio en un mismo marco estratégico. A través de esta combinación, el régimen persigue alterar percepciones públicas, debilitar la cohesión interna de sus adversarios y proyectar una imagen de fortaleza y legitimidad internacional, empleando capacidades tecnológicas y estructuras de carácter clandestino (Microsoft, 2023b, 2024; Sha, 2025; CyberProof, 2025). Esta aproximación refleja un modelo de influencia estatal sofisticado que opera en la convergencia entre el ciberespacio y la información pública, difuminando las fronteras entre la diplomacia, la inteligencia y la propaganda (Lee, 2025; England, 2025).

Operativa APT en las campañas de influencia

Las operaciones de desinformación deben comprenderse como procesos interconectados en los que confluyen dimensiones técnicas y narrativas. Los grupos APT y los operadores de influencia (IO) forman parte de un mismo ecosistema estratégico, donde los primeros aportan las capacidades ofensivas y la infraestructura tecnológica, mientras que los segundos diseñan, adaptan y difunden los relatos. Esta interacción, que puede resumirse en la secuencia "*Hack → Leak → Amplify*", ha llevado a definir estas dinámicas como campañas híbridas, consolidándolas como una herramienta clave dentro de las operaciones de influencia en el ciberespacio (European Union External Action, 2025). En determinados casos, los grupos APT operan como actores híbridos capaces de combinar la intrusión técnica con la manipulación informativa. En otros, la estructura es más compartimentada distinguiendo entre el actor responsable del compromiso de la infraestructura y aquel que explota narrativamente los resultados obtenidos. En ambos

escenarios, los actores APT constituyen un eslabón crítico al proporcionar los activos esenciales (datos, credenciales y documentos internos) y los recursos tecnológicos necesarios para la continuidad de las operaciones (NATO Strategic Communications Centre of Excellence, 2021).

Estos actores pueden intervenir en distintos niveles de la campaña. Actúan como promotores o patrocinadores cuando representan directamente intereses estatales; como productores y recolectores de contenido al extraer datos de sistemas comprometidos; como amplificadores mediante *bots* que difunden las narrativas; o como facilitadores al proveer infraestructura técnica compartida bajo modelos similares al *infrastructure-as-a-service*. En algunos casos recurren a dominios comprometidos o servidores de comando y control (C2) como nodos de diseminación, o despliegan contra-narrativas destinadas a obstaculizar la atribución (Cordey, S., 2019).

La incorporación de tecnologías emergentes, especialmente los *deepfakes*, ha multiplicado el alcance y la sofisticación de estas campañas. Entre 2023 y 2024 se registraron más de 82 casos en 38 países dirigidos a figuras públicas, incluyendo políticos, lideres estatales y otras personalidades, con fines de manipulación, difusión de declaraciones falsas, injerencia electoral o difamación. Estas prácticas incluyen la elaboración de *deepfakes* mediante IA, así como el uso de medios sintéticos y otros contenidos generados por IA destinados a fabricar declaraciones o materiales falsos que influyen o desacreditan a los objetivos (Insikt Group, 2024b).

Finalmente, la infraestructura técnica constituye un elemento decisivo. Los actores APT proporcionan dominios, servidores C2, cuentas comprometidas y servicios web que sustentan tanto el acceso inicial como la diseminación de contenidos. Con frecuencia, dicha infraestructura se gestiona bajo esquemas de externalización o compartición de recursos, lo que no solo complica los procesos de atribución, sino que también refuerza la resiliencia operativa. Esta doble función, control de la infraestructura y manipulación de la información, convierte a los APT en actores imprescindibles dentro de las campañas de influencia estatales, al situarlos en el punto de convergencia entre capacidad técnica y la estrategia comunicativa (Microsoft, 2024a).

Hacktivismo

El concepto de *hacktivismo* ha cambiado de forma significativa en los últimos años. Tradicionalmente, el término hacía referencia a colectivos o individuos que utilizaban medios digitales para reivindicar causas sociales, políticas o ideológicas. Estos hacktivistas clásicos podían atacar páginas web, filtrar información o lanzar mensajes de protesta, pero su actuación estaba vinculada a la defensa de unos valores concretos, en muchos casos alineados con movimientos sociales más amplios. Sin embargo, en la actualidad se observa un fenómeno distinto, conocido como "faketivismo" o *hacktivismo fake* (ThreatConnect, 2016). Este nuevo tipo de hacktivismo se caracteriza por ser más agresivo y desvirtuado del espíritu original, actuando menos como forma de protesta social y más como una herramienta de influencia geopolítica. En este contexto, el hacktivismo actual no solo libra una batalla técnica mediante ciberataques, sino que participa en una batalla dialéctica destinada a influir en la opinión pública. Se trata de una disputa narrativa donde se intenta legitimar a un bando y deslegitimar al contrario, reforzando el marco propagandístico que acompaña a la confrontación militar.

Un ejemplo claro de esta transformación se produjo con el inicio de la guerra en Ucrania, cuando surgieron o se consolidaron grupos de *hacktivismo* instrumentalizado por estados. Entre ellos destaca NoName057(16), un colectivo que ha adquirido notoriedad internacional (Intel471, 2025). Estos actores ya no se presentan como defensores de valores universales, sino como partidarios explícitos de un país concreto, convirtiéndose en una extensión digital de los conflictos bélicos tradicionales. Por ello, a la hora de identificar actores hacktivistas en campañas de desinformación, resulta esencial comenzar por el análisis de su motivación o propósito. Solo así es posible diferenciar entre quienes actúan como activistas digitales clásicos, con causas definidas y autonomía relativa, y quienes representan un *hacktivismo* "de estado", donde el objetivo principal es inclinar la balanza de la opinión pública internacional en favor de los intereses de una nación.

Los llamados hacktivistas "de estado" se distinguen por su forma de operar y por la elección de sus objetivos. A diferencia de los hacktivistas clásicos, estos grupos concentran su actividad en molestar e interrumpir servicios esenciales mediante ataques de denegación de servicio distribuida (DDoS). Sus blancos suelen ser instituciones gubernamentales, empresas estratégicas o industrias críticas asociadas a un país concreto, todo ello dentro de campañas coordinadas que buscan erosionar la imagen y credibilidad de un adversario.

¿Cómo funciona a alto nivel una campaña de desinformación hacktivista?

Las campañas de desinformación protagonizadas por grupos hacktivistas suelen seguir un patrón repetitivo, en el que la acción técnica se combina con una narrativa diseñada para maximizar el impacto. Este ciclo puede dividirse en cinco fases principales:

1. **Anticipación y selección de objetivos.** Los grupos monitorizan el contexto internacional en busca de eventos de alto valor simbólico: elecciones, cumbres internacionales, conflictos bélicos o decisiones judiciales, entre otros, y, a partir de ellos, determinan los objetivos vinculados, como por ejemplo páginas de partidos políticos, ministerios, medios de comunicación, instituciones electorales o empresas estratégicas. Esta fase está marcada por un fuerte componente ideológico y propagandístico, donde la elección del objetivo refuerza la narrativa del grupo.

2. **Preparación y coordinación.** A través de canales públicos en X y Telegram, los actores anuncian campañas o buscan simpatizantes que se sumen a los ataques, a menudo construyendo alianzas con otros para aumentar el impacto y difundir sus mensajes para generar un daño reputacional. En paralelo, se trasladan a canales privados o cifrados para planificar detalles técnicos, repartir herramientas (DDoS *kits, scripts,* credenciales) y definir los mensajes de propaganda que acompañarán la operación. La coordinación incluye la preparación de contenido visual o textual (memes, vídeos, comunicados) que se lanzará de forma simultánea al ataque.

3. **Ejecución técnica del ataque**. La ofensiva se materializa normalmente mediante ataques DDoS contra páginas web gubernamentales o corporativas, aunque en algunos casos con acciones adicionales como *defacements* (cambiar la apariencia de un sitio web) o filtraciones de datos para ganar un mayor impacto. Pese a su carácter normalmente efímero, el efecto técnico suele bastar para generar titulares y provocar fricción mediática.

4. **Difusión y propaganda**. Una vez lanzado el ataque, los actores utilizan X, Telegram y otras redes sociales para reivindicar la autoría, exagerar los efectos y ridiculizar al adversario. Aquí entran en juego la sátira, los memes y, cada vez más, contenido manipulado con inteligencia artificial para aumentar la viralidad. El objetivo es que la narrativa del grupo alcance a medios de comunicación y sea replicada por terceros, amplificando su alcance.

5. **Explotación narrativa y legitimación**. Tras la difusión, se inicia un proceso de justificación del ataque donde se presenta como una respuesta "necesaria" ante una injusticia o como un golpe simbólico contra un enemigo político. Esto refuerza la idea de que el grupo defiende una causa legítima y que su acción trasciende lo técnico, insertándose en la batalla dialéctica del conflicto. En algunos casos, el ataque se acompaña de información falsa o manipulada (por ejemplo, noticias falsas sobre un supuesto fallo electoral) para legitimar la acción técnica y aumentar la confusión.

Además del impacto técnico, que a menudo es limitado y temporal, el verdadero valor de estas operaciones radica en la dimensión propagandística. Para ello, los hacktivistas de estado explotan de manera sistemática las plataformas de comunicación abiertas y cerradas.

- X (antes Twitter) y Telegram (canales públicos) se utilizan como altavoces para dar a conocer sus ataques, reivindicar la autoría y proyectar una imagen de fuerza. En estos espacios también buscan atraer la atención mediática y captar nuevos simpatizantes.

- Una vez establecida la narrativa y generada cierta atracción, los grupos suelen migrar a canales privados en Telegram u otras plataformas de mensajería cifrada, donde planifican y coordinan los ataques con mayor discreción. Estos espacios restringidos se convierten en el núcleo operativo de la organización, permitiendo compartir herramientas, coordinar objetivos y establecer jerarquías internas.

El uso de Telegram como plataforma principal de coordinación y difusión de los hacktivistas "de estado" se ha visto recientemente limitado, ya que muchos de sus grupos están siendo bloqueados o dejados inactivos. Esto genera un reto adicional: la pérdida de histórico y trazabilidad de las operaciones si no se retienen evidencias de manera constante. De ahí la importancia de establecer mecanismos de monitorización y archivado frecuente, que permitan conservar un registro de los ataques y narrativas utilizadas por estos actores.

Este modelo híbrido de comunicación —público para la propaganda, privado para la coordinación— ha demostrado ser altamente eficaz. Por un lado, amplifica el impacto psicológico de operaciones de bajo coste; por otro, dificulta la labor de monitorización e infiltración, ya que la transición entre lo abierto y lo cerrado fragmenta el rastro digital de los actores.

En consecuencia, el seguimiento y la identificación de hacktivistas de estado requiere una metodología que combine el análisis de sus narrativas públicas con el rastreo técnico de sus actividades en entornos privados, permitiendo así entender la conexión entre sus fines propagandísticos y sus acciones técnicas en el ciberespacio.

En sus mensajes, estos grupos no se limitan a anunciar ataques. Emplean un lenguaje cargado de sátira y burla, buscando ridiculizar a sus adversarios y generar mayor viralidad en redes sociales. Además, se apoyan en la inteligencia artificial para crear contenido audiovisual: memes y vídeos manipulados donde políticos aparecen realizando acciones contradictorias a sus valores o discursos. Esta combinación de humor, falsificación y manipulación persigue erosionar la reputación pública de figuras clave y reforzar la narrativa que legitima sus ataques.

Frente a estas dinámicas, la anticipación se convierte en la herramienta principal para identificar campañas de desinformación impulsadas por hacktivistas. El análisis prospectivo de eventos internacionales, conflictos bélicos y tensiones políticas permite prever cuáles serán los objetivos más probables. Un ejemplo claro es el de los procesos electorales: durante unos comicios, los hacktivistas "de estado" suelen dirigir sus ataques contra las páginas web de partidos políticos, organismos electorales o medios de comunicación asociados. A los ataques técnicos, generalmente DDoS, se suma la difusión de información falsa diseñada para generar dudas sobre la legitimidad de las elecciones. De este modo, logran amplificar el impacto del ataque técnico con una campaña narrativa que busca desestabilizar y polarizar la opinión pública.

En definitiva, la identificación de operaciones de desinformación con participación de hacktivistas requiere una metodología basada en la monitorización continua, la preservación de evidencias digitales y la anticipación de escenarios de riesgo, especialmente en torno a procesos políticos sensibles y contextos de alta tensión internacional.

A continuación, se presentan algunos de los actores hacktivistas más activos y representativos a nivel global. Conviene tener en cuenta que el panorama es altamente dinámico: surgen nuevos grupos con rapidez, otros desaparecen o se transforman, y las narrativas y objetivos pueden variar en cuestión de semanas. Por ello, esta selección refleja una fotografía del momento actual más que un mapa estable y definitivo.

- NoName057(16) es un grupo *hacktivista* prorruso activo desde marzo de 2022. Atacan principalmente a instituciones públicas, empresas y servicios de países que apoyan a Ucrania y a la OTAN. Su narrativa se centra en "defender la verdad" y restaurar la justicia en favor de Rusia. Sus ataques son sobre todo DDoS. En ocasiones, sus campañas se desarrollan en paralelo con los de otros colectivos

hacktivistas prorrusos, reflejando alianzas fluidas y tácticas compartidas dentro del ecosistema de *hacktivismo* pro-ruso (Intel471, 2025, SOCRadar, 2023, Cyble, 2025).

- Mr Hamza es un actor activo desde 2024, probablemente de origen marroquí. Defiende narrativas pro-Palestina y anti-Israel, combinando activismo político con actividades comerciales. Sus ataques son principalmente DDoS, junto con *defacement* y filtraciones de información. Ataca sobre todo a Israel, Estados Unidos, India y diversos países europeos, centrándose en sectores gubernamentales, defensa, telecomunicaciones y finanzas. Durante los momentos de mayor tensión geopolítica, su actividad se intensifica y, en ocasiones, se alinea con otros colectivos hacktivistas de motivación similar, lo que amplifica su alcance operativo (SOCRadar, 2025; Meidan, 2025b).

- BL4CK CYB3R es un grupo hacktivista pro-camboyano, activo en Telegram desde marzo de 2025. No se centran en un solo sector, pero suelen priorizar sitios web gubernamentales y entidades financieras como objetivos principales en múltiples países. Sus acciones buscan visibilidad política más que beneficios económicos (GroupIB, 2025).

- Keymous+ es un grupo surgido a finales de 2023, con más de 700 ataques. Aunque se presentan como hacktivistas norteafricanos, carecen de una narrativa clara y atacan objetivos muy variados tanto geográficamente como a nivel de sectores, si bien las instituciones gubernamentales son las más atacadas. Operan dos equipos: uno de filtraciones inactivo y otro de DDoS muy activo. Su actividad muestra también rasgos de DDoS-as-a-Service, ligados al servicio EliteStress (Meidan, 2025a).

Finalmente, en el caso de España, destaca la existencia de un canal de Telegram denominado *Desinformador ruso*, donde se difunde un discurso claramente prorruso y despectivo hacia las instituciones españolas. Este espacio no solo replica narrativas alineadas con la propaganda del Kremlin, sino que además amplifica el eco de los ataques reivindicados por grupos hacktivistas, ensalzando sus campañas y reforzando la percepción de legitimidad de sus acciones en la opinión pública (Colás y Rojas, 2025).

En conclusión, el *hacktivismo* ha pasado de ser una expresión digital de protesta social a convertirse en un instrumento estratégico dentro de la confrontación geopolítica. Este tránsito hacia el "faketivismo" refleja un uso cada vez más instrumentalizado y agresivo, en el que la narrativa pesa tanto o más que el efecto técnico de los ataques. La combinación de operaciones de bajo coste —principalmente DDoS— con campañas propagandísticas

en redes sociales y el uso de contenidos manipulados demuestra que el verdadero valor de estas acciones radica en su capacidad para polarizar y erosionar la confianza pública.

La identificación y clasificación de estos actores exige, por tanto, una metodología que combine la monitorización continua, la preservación de evidencias digitales y la anticipación de escenarios sensibles, como elecciones o conflictos internacionales. Solo de este modo es posible diferenciar a los hacktivistas tradicionales, con causas autónomas y limitadas, de aquellos que actúan como extensiones digitales de los estados en disputa. El ecosistema *hacktivista* es altamente volátil: surgen nuevos grupos, otros desaparecen o mutan, y las narrativas cambian en cuestión de semanas. Por ello, más que un mapa estable, el panorama actual debe entenderse como una fotografía en constante movimiento. En este contexto, el análisis crítico de motivaciones, objetivos y narrativas resulta esencial para comprender cómo estos actores encajan en las operaciones de desinformación y cuál es su papel en la batalla por la opinión pública global.

Cibercrimen

Si bien el término cibercrimen cobró mayor relevancia durante las décadas de 1980 y 1990, coincidiendo con el auge de los virus informáticos y los primeros accesos no autorizados a sistemas, la expansión global de Internet a partir del año 2000 permitió que este fenómeno evolucionara de acciones simples a operaciones cada vez más complejas.

En la actualidad, el cibercrimen constituye una de las principales amenazas de la era digital, afectando de forma directa a compañías, gobiernos y ciudadanía. Su constante crecimiento, adaptación y sofisticación técnica lo convierten en un desafío permanente para los investigadores privados y las Fuerzas y Cuerpos de Seguridad del Estado.

La desinformación, entendida como información verificablemente falsa o engañosa que se crea, presenta y divulga con fines lucrativos o para engañar deliberadamente a la población, ha sido tradicionalmente asociada a campañas impulsadas por actores estatales extranjeros en contextos de guerra híbrida o influencia geopolítica. No obstante, con el paso del tiempo, actores del cibercrimen que operan en foros *underground* y aplicaciones de mensajería cifrada también han adoptado el uso de narrativas manipuladas y estrategias informativas para la consecución de sus objetivos, evidenciando un creciente interés por la desinformación como herramienta operativa.

Desinformación como engaño o fraude digital

Uno de los usos más recurrentes de la desinformación, por parte de actores del ecosistema del cibercrimen, consiste en su empleo como instrumento de engaño o fraude, mediante el cual se explota la confianza o el desconocimiento de las víctimas con el propósito de influir en su conducta (Hu et al., 2024). A través de campañas cuidadosamente diseñadas para inducir miedo, confusión o sensación de urgencia, los atacantes logran manipular el comportamiento de individuos u organizaciones con el fin de obtener información confidencial, acceso privilegiado o beneficio económico. Esto puede observarse en el uso de llamadas telefónicas, mensajes SMS o correos electrónicos fraudulentos que, suplantando la identidad de autoridades, entidades o instituciones, generan confianza en la víctima como para que esta facilite su información personal, realice pagos o descargue archivos maliciosos. Del mismo modo, dado el elevado número de usuarios que las utilizan, los actores del cibercrimen se apoyan en las redes sociales para maximizar el alcance de sus campañas, consolidándolas como una herramienta clave en su difusión de desinformación.

Un ejemplo representativo de ello se observó durante la pandemia de la COVID-19, cuando se registró un incremento significativo de campañas de phishing, acompañadas de narrativas falsas relacionadas con vacunas, certificados médicos y ayudas gubernamentales (Europol, 2020). Estas acciones no solo explotaron la incertidumbre generalizada de la población, sino que además alcanzaron un elevado índice de éxito al aprovechar un contexto de crisis sanitaria global, en el que la urgencia informativa y el miedo facilitaron la manipulación y el engaño. A ello se sumó la amplificación masiva del contenido desinformativo, a través de redes sociales y aplicaciones de mensajería instantánea, donde la rápida propagación de los rumores dificultó la detección y la verificación, lo que permitió a los atacantes adaptar sus mensajes en tiempo real.

Desinformación con fines de extorsión y daño reputacional

La desinformación también es utilizada por el cibercrimen como un instrumento de coacción frente a individuos, empresas e instituciones, bajo la amenaza de difundir información veraz o manipulada que pueda perjudicar su imagen y, en consecuencia, socavar su reputación frente a terceros (Kalajdziovski, y Collard, 2025). Este tipo de práctica ha adquirido una relevancia creciente en los últimos años, en un contexto en el que la exposición digital y la confianza pública se han convertido en factores determinantes. Estas dinámicas combinan elementos propios de los incidentes de ciberseguridad, como el *ransomware* o la filtración selectiva de información, con

estrategias psicológicas y comunicativas diseñadas para influir en la opinión pública y generar presión sobre la víctima. Por ello, la reputación digital se configura como un activo crítico y altamente vulnerable, cuyo deterioro puede tener consecuencias en términos de legitimidad institucional, relaciones comerciales o estabilidad organizacional. Además, en la actualidad se caracteriza por la inmediatez y la viralidad informativa, lo que facilita que la difusión de un contenido falso pueda propagarse rápidamente, dificultando la respuesta ante el daño reputacional.

Por otro lado, los actores del cibercrimen también han llegado a emplear la desinformación contra competidores, llevando a cabo campañas de desprestigio en un mercado criminal altamente competitivo. Entre las tácticas más habituales destacan las acusaciones falsas de fraude, colaboración con las autoridades o incluso infiltración como agentes encubiertos, todas ellas orientadas a influir en alianzas, desestabilizar a rivales y controlar la narrativa del ecosistema delictivo. Este uso estratégico de la desinformación refleja la creciente convergencia entre los ámbitos técnico y cognitivo del ciberespacio, donde las operaciones trascienden el sabotaje o el robo de información para incidir directamente en la percepción, la credibilidad y la confianza de los actores involucrados, ampliando así el alcance y la complejidad de las ciberamenazas.

Desinformación para la obtención de reputación

Más allá del beneficio económico, los actores del cibercrimen también han empleado la desinformación con fines reputacionales dentro del propio ecosistema delictivo en el que operan. En estos entornos cerrados y altamente competitivos, la percepción de poder, la demostración de capacidades técnicas y el acceso exclusivo a información sensible se configuran como factores determinantes para obtener legitimidad, acceder a transacciones ilícitas de mayor valor y generar confianza frente a terceros. En este contexto, la atribución falsa de intrusiones o compromisos a empresas, instituciones públicas o entidades gubernamentales se utiliza de manera deliberada para incrementar la reputación digital y simular una mayor capacidad operativa (Check Point, 2025). De igual modo, la difusión de supuestas muestras o bases de datos robadas, elaboradas en realidad a partir de información falsificada o generadas mediante el uso de inteligencia artificial, se ha convertido en una práctica cada vez más habitual dentro del ámbito criminal, que permite atraer la atención e interés de posibles compradores o colaboradores.

Estas estrategias, sustentadas principalmente en la manipulación informativa, buscan incrementar la visibilidad y credibilidad de los actores ante su entorno, generando un efecto de autoridad y confianza que facilita nuevas oportunidades delictivas y fortalece su posicionamiento criminal.

Desinformación como medida de contrainteligencia

Dada la creciente vigilancia e infiltración en foros *underground* por parte de investigadores especializados y de las Fuerzas y Cuerpos de Seguridad del Estado, numerosos actores han implementado mecanismos de contrainteligencia basados en la desinformación, con el propósito de dificultar su atribución, sembrar dudas y desviar la atención de las autoridades. Una de las tácticas más habituales consiste en la atribución falsa de ciberataques a actores rivales, con el objetivo de confundir y desorientar a las autoridades. En otras ocasiones, los cibercriminales difunden rumores falsos sobre su propia detención, fallecimiento o retirada de la actividad, buscando reducir la presión y el seguimiento de sus operaciones (Franceschi-Bicchierai, 2016).

Estas prácticas reflejan una evolución cualitativa del cibercrimen, que ha pasado de un conjunto de acciones aisladas a una estructura organizada, adaptativa y consciente del entorno que la rodea. Lejos de responder a una lógica improvisada, estas campañas de desinformación siguen patrones estructurados que abarcan desde la planificación estratégica y la definición de objetivos, hasta la selección de audiencias, el desarrollo narrativo y la elección de canales de difusión.

REDES DE INFLUENCIA TECNOLÓGICA

La convergencia entre las actividades de diversos actores de amenaza, la desinformación y el uso intensivo de tecnologías basadas en inteligencia artificial ha dado lugar a un entramado de redes de influencia tecnológica que pone a prueba los mecanismos tradicionales de seguridad, atribución y gobernanza. Este fenómeno se evidencia en la consolidación de ecosistemas como el *Disinformation as a Service (DaaS)* y el *Crime as a Service (CaaS)*, en los que la exfiltración de datos personales, la automatización de campañas de manipulación informativa y la utilización de nuevas vías de financiación fortalecen la conexión entre actores estatales y no estatales. En este escenario, la protección de la privacidad y la trazabilidad de los datos adquieren una relevancia estratégica, no solo como un derecho esencial, sino como una responsabilidad compartida orientada a preservar la confianza pública, la integridad institucional y la resiliencia democrática frente a los riesgos derivados de la manipulación digital y el uso instrumental de la tecnología.

Identificación de nodos tecnológicos clave

En el contexto actual estamos presenciando una evolución del conjunto de elementos utilizados en operaciones digitales en el contexto de la desinformación y el fenómeno de la injerencia y manipulación extranjera de la información (FIMI). La evolución de técnicas de IA ha ensanchado el espectro de capacidades de la desinformación, consolidando lo que podemos venir a tipificar como ecosistema *DaaS –Disinformation as a Service–* (Departamento de Seguridad Nacional, 2024; DarkOwl, 2024). Además, es preciso poner de relieve el trasvase de operaciones e intereses entre el ámbito del cibercrimen y el dominio específico de acciones de interferencia esponsorizadas por estados (T-Sanct Technologies, 2025). Todo ello supone un conjunto de desafíos adicionales a la hora de llevar a término procedimientos de (ciber)atribución efectivos y eficaces. Sin que ello suponga una aproximación taxonómica, a continuación, se resumirán algunas de las amenazas asociadas al aumento de ataques de exfiltración de datos personales y sus implicaciones en suplantación de actores e instituciones de relevancia, al progresivo uso de IA en campañas de desinformación, así como a la aparición de nuevas vías de financiación para soportar actividades coordinadas en ataques reputacionales y operaciones de influencia.

La protección de datos personales como vehículo para la lucha contra la desinformación

La protección de la privacidad se ha venido considerando de forma tradicional como un elemento dentro del ámbito del derecho a la intimidad. La complejidad inherente al concepto de privacidad se ha visto amplificado especialmente en la última década como consecuencia de la popularización de redes sociales y la mediatización de nuestra realidad social y política. Al grueso de modalidades a través de las cuales exponemos y ponemos en riesgo nuestra privacidad, hay que añadir el incremento en el grado de sofisticación y accesibilidad a herramientas basadas en inteligencia artificial, las cuales facilitan tanto la captura de información como su instrumentalización para obtener beneficios de datos personales, o para poner en práctica dinámicas que pueden suponer un riesgo para los ciudadanos a los que están asociados dichos datos (Adler et al., 2024).

En el contexto actual, las brechas de seguridad que llevan asociadas la perdida de datos personales han de ser consideradas como indicio de potenciales campañas FIMI. En este sentido, es preciso dar cuenta del trasvase de competencias entre el ecosistema APT y el ámbito FIMI, dado el creciente solapamiento entre técnicas propias del cibercrimen y procesos de manipulación. La información obtenida a través de ataques de ingeniería social básicos, así como de tácticas *ransomware* o de *infostealers* pueden constituir la base de conocimiento de movimientos de segunda fase en las primeras etapas de la *Kill Chain* asociadas en este caso a APT (Privacy International, 2021; The Economist, 2025a, 2025b), pudiendo habilitar también dinámicas más complejas de manipulación persistente como APM –*Advanced Persistent Manipulation*– (Rachel James, 2024; Hamayun, 2025), una forma de ataque que utiliza inteligencia artificial para sostener campañas de manipulación social continuadas y adaptadas a cada objetivo.

La coordinación de actividades con foco en erosionar la posición de marca, la reputación y, en definitiva, la confianza de individuos, organizaciones e instituciones es de altísima relevancia, de forma que la evaluación de este tipo de riesgo debe ser tenida muy en cuenta. La raíz misma de este problema es la que motiva los esfuerzos actuales por efectuar análisis en profundidad de los tipificados como riesgos ESG –*Environment, Sustainability and Governance*– (Mishra, 2025). Aquí conviene recordar que, dentro de las diversas aproximaciones conceptuales, la confianza se debe interpretar en términos de competencia, benevolencia e integridad del potencial actor confiable (Degli-Esposti, & Arroyo, 2021). Este cuidado, esta cura sui de lo institucional abarca, mejor dicho, ha de abrazar a lo instituyente. En efecto, el contexto social y político que engloba a España y a Europa está cada vez más mediado por el ciberespacio, por la interacción y con-formación del sujeto con lo digital. El leitmotiv "los datos son el nuevo oro" es engañoso: los datos es la estructura misma de nuestro día, y son oro fruto y resultado

de su uso, de su explotación, de su aplicación. Si tal uso o aplicación cae en manos de un tercero, es bastante plausible que el rédito de tales datos para nosotros se convierta más en un debe, un menoscabo.

Operaciones como las de *Cambridge Analytica*, pero también todas las campañas de *phishing* y de ingeniería social, ponen de relieve el riesgo asociado a la huella digital que dejamos fruto de nuestra actividad en el ciberespacio. Esa huella digital puede ser, y de hecho es, utilizada por parte de actores de injerencia y manipulación con especial capacidad e interés en dividir nuestra sociedad aprovechando nuestros sesgos, nuestras vulnerabilidades psicológicas y cognitivas (Departamento de Seguridad Nacional, 2023).

Como bien queda referida en la conceptualización del "*dataself*", si el dato es tan preciado, hemos de cuidarlo y protegerlo (Lupton, 2019). En este punto, de nuevo, es preciso realizar un giro conceptual. Si solemos afirmar que una sociedad con una democracia sólida es resistente a ataques externos y operaciones de influencia perpetradas por terceros, hemos de tener muy presente que una democracia es tan sólida como la suma colectiva que emana de ciudadanos con una cultura digital bien acendrada, con unos buenos hábitos digitales y con una higiene digital que muestra inermes intentos de *phishing*, ingeniería social y de intoxicación informativa mediante contenido creado artificialmente, sin contexto y sin una base de evidencia trazable y verificable.

Aplicación de IA en cibercrimen y en FIMI

La evolución de los diversos ciberataques en el ecosistema CaaS (*Crime as a Service*) ha deparado un creciente protagonismo de las brechas de privacidad y exfiltración de datos. En efecto, en el último año el rol preponderante del *ransomware* en el dominio de los ciberataques ha quedado relegado por el auge de los denominados infostealers,. Un claro ejemplo de ello es la desarticulación de *Lumma Stealer* (Masada, 2025). El robo de datos y su comercio constituye una de las actividades con más rédito en CaaS. Este rédito, por otro lado, debe ponerse en el contexto de la alta rentabilidad que el cibercrimen supone no sólo para grupos y organizaciones criminales, sino también para Estados. El estudio en profundidad de las TTP asociadas al mercado de datos en la Dark Web es de gran importancia (Marjanov, y Hutchings, 2025), por ejemplo, para poder diseñar estrategias adecuadas de protección frente a la automatización de ataques de suplantación mediante el uso de datos robados de otros usuarios a través de aplicaciones como Linken o Fraudfox, que permiten recrear entornos digitales legítimos y manipular la huella digital de navegación para eludir los sistemas de detección y verificación antifraude.

Dentro del ecosistema del *ransomware*, además de la tendencia creciente a remplazar el costoso despliegue de soluciones para el cifrado de datos de las víctimas por la

"simple" extorsión mediante publicación parcial de datos (McIntosh et al., 2024), es especialmente significativo el giro de grupos criminales como *LunaLock*, que realizan la extorsión mediante la amenaza de enviar los datos sustraídos a empresas de IA para que puedan emplearlos de cara a entrenar sus modelos. De esta forma, el ecosistema RaaS (*Ransomware as a Service*) aprovecha el creciente temor de la industria de creadores de contenido frente a la capacidad de imitación de la IA.

Por último, y como actualización de los ángulos cubiertos en lo relativo al uso de IA en el dominio de la desinformación en el capítulo 5 de los trabajos de 2023 del Foro contra las campañas de desinformación, es preciso dar cuenta de la modificación e inclusión de nuevas TTP fruto del uso de IA generativa por parte de actores de amenazas. Informes como el publicado por Microsoft constituyen una llamada de atención para poner el foco en el ecosistema de servicios y microservicios de IA generativa (Microsoft, 2024c).

Nuevas vías de financiación de actividad FIMI

El fenómeno TikTok (Annabell et al., 2025), pero también las plataformas de *gaming* (Levi, 2009) o los servicios como tarjetas regalo, los programas de fidelización y de millas, deparan un nuevo canal de financiación que escapa a los controles *KYC* y *AML (FinCrime Central, 2024)*. En el caso de las grandes plataformas o *Very Large Online Platforms* (*VLOPs*), la monetización de contenido, su moderación y su trazabilidad han de amoldarse a lo que establece el Reglamento de Servicios Digitales (DSA). Es más, el artículo 40 de la DSA establece la obligación de las *VLOPs* de proporcionar acceso de forma transparente a la información relativa a la aplicación de políticas de moderación de contenido (Trujillo et al., 2025).

En efecto, el trabajo colaborativo entre comunicadores y expertos en el análisis de fuentes abiertas de datos puede y debe contribuir a mejorar las medidas para implementar KYC/AML y evitar la proliferación de contenido tóxico, campañas de manipulación o estrategias de radicalización en redes sociales o plataformas, como las de *gaming*.

Trazabilidad y análisis de criptomonedas

Las criptomonedas son un tipo de moneda digital que emplea la criptografía para garantizar la seguridad de las transacciones y el control de la creación de nuevas unidades. Estas monedas operan sobre una tecnología denominada *blockchain* (cadena de bloques), que es básicamente una base de datos descentralizada que registra todas las transacciones generadas en la red.

Blockchain funciona como un libro de contabilidad, generalmente público, en el cual cada "página" representa un bloque, y la conexión entre ellos sería la cadena. Su carácter descentralizado implica que la información no se almacena en un único servidor, como sucede en las instituciones financieras tradicionales, sino que se distribuye entre numerosos ordenadores a nivel global. De este modo, los registros pueden consultarse de forma gratuita y confiable mediante los denominados *exploradores de bloques*. Toda la información registrada en la cadena de bloques se conoce como información *on-chain*, e incluye los datos de las direcciones con sus respectivos saldos, así como las transacciones realizadas por los participantes de la red.

La primera y más reconocida criptomoneda es Bitcoin, creada en 2009 por una entidad o grupo de personas bajo el seudónimo Satoshi Nakamoto. Bitcoin suele compararse con el oro debido a su oferta limitada, solo existirán 21 millones de unidades, y su uso como valor refugio. Como todas las criptomonedas, utiliza la tecnología de *blockchain* para realizar transacciones seguras y descentralizadas. Cada transacción de Bitcoin es registrada en su propia *blockchain*, lo que permite la trazabilidad y transparencia de sus operaciones.

Las direcciones de cambio en Bitcoin son direcciones a las que los Bitcoins retornan como "cambio" después de una transacción. Se puede comparar a cuando se paga con un billete de 20€ por algo que cuesta 15€. La vuelta de 5€, que es el "cambio", se guarda en un bolsillo "diferente". Con Bitcoin, esto ocurre digitalmente, el cambio suele ir a una dirección de Bitcoin diferente a la originaria, pero perteneciente al mismo usuario inicial.

Ethereum es otra criptomoneda importante, pero va más allá del simple concepto de moneda digital de Bitcoin. Su infraestructura tecnológica permite la creación y ejecución de contratos inteligentes (*smart contracts*), es decir, programas informáticos que se ejecutan automáticamente cuando se cumplen determinadas condiciones previamente establecidas.

Esta funcionalidad amplía de forma considerable las posibilidades de aplicación de la tecnología *blockchain*, haciendo posible el desarrollo de *sistemas financieros descentralizados* (*DeFi*), y/u *Organizaciones Autónomas Descentralizadas* (*DAO*).

Las finanzas descentralizadas o DeFi representan un nuevo paradigma financiero que opera sobre redes *blockchain* y prescinde de intermediarios tradicionales, como bancos o entidades FinTech. Los servicios DeFi permiten realizar transacciones de carácter global (*borderless*) y sin validación, autorización o KYC (*permissionless*), lo que significa que cualquier persona, en cualquier lugar del mundo, puede participar sin necesidad de un permiso previo.

Por su parte, las DAO son estructuras organizativas que funcionan mediante *smart contracts*. Estas entidades pueden entenderse como empresas que operan en una *blockchain*, sin una estructura de liderazgo centralizada, y cuyas decisiones son tomadas por los titulares de sus tokens (se habla pues de gobernanza descentralizada).

Actualmente, existen numerosas redes *blockchain*, siendo la mayoría pequeñas modificaciones de las redes principales Bitcoin y Ethereum. Un ejemplo de ello es la red Tron, que replica la arquitectura de Ethereum. Debido a su menor nivel de uso y demanda, Tron ofrece transacciones significativamente más económicas, al tiempo que permite el funcionamiento de criptoactivos equivalentes, como la *stablecoin* USDT emitida en su propia red.

Por último, el término criptoactivo engloba a todas las formas de activos digitales basados en tecnología *blockchain*, incluyendo las criptomonedas, los tokens y los *Non-Fungible Tokens (NFT*). Dentro de este conjunto, las *stablecoins* ocupan un lugar particular, ya que su valor se mantiene vinculado a un activo estable, como el dólar estadounidense, lo que les permite evitar las fluctuaciones de precio propias de otras criptomonedas. Gracias a esta estabilidad, las *stablecoins* se utilizan ampliamente en los ecosistemas DeFi como medio de intercambio y reserva de valor.

Transacciones y Exchanges

Las criptomonedas se custodian en billeteras digitales (*wallets*), las cuales permiten enviar y recibir activos digitales de manera descentralizada a cualquier parte del mundo, casi de forma instantánea y con costos de transacción mínimos, en muchos casos apenas unos céntimos de euro o incluso sin comisión significativa. Cada transacción realizada con criptomonedas queda registrada en la *blockchain* correspondiente. Por ejemplo, una transferencia de bitcoins se inscribe en la cadena de bloques de Bitcoin, lo que posibilita el seguimiento del historial completo de operaciones efectuadas en dicha red. No obstante, la información almacenada en la *blockchain* es de carácter técnico y limitado. Esta suele incluir los datos esenciales de la operación, como las direcciones de origen y destino, el tipo y la cantidad de criptoactivo transferido, la fecha y hora exacta de la transacción, así como su hash (identificador único que garantiza su autenticidad e inmutabilidad).

Además, existe otro tipo de transacciones que se realizan a través de las denominadas casas de cambio o *exchanges* de criptomonedas. Estas plataformas, generalmente centralizadas, permiten a los usuarios realizar diferentes operaciones (comprar, vender o intercambiar criptomonedas), ya sea por otras criptomonedas o por monedas tradicionales (por ejemplo, euro o dólar). Entre los *exchanges* más conocidos se encuentran Coinbase, Kraken, Binance o Bit2Me.

Existen dos categorías principales de *exchanges* de criptomonedas. La primera corresponde a los *Centralized Exchanges* (*CEXes*), los cuales operan de manera similar a las instituciones financieras tradicionales, ya que son gestionados por una entidad que actúa como intermediaria. Estos *exchanges* suelen requerir a sus clientes un proceso de verificación de identidad conocido como *Know Your Customer* (*KYC*), en cumplimiento con las normativas internacionales de prevención de blanqueo de capitales y financiación del terrorismo (AML/CTF).

En los CEX, muchas de las operaciones internas, como los intercambios de criptoactivos entre usuarios, se realizan de manera *off-chain*, es decir, las transacciones ocurren 'fuera' de la *blockchain* y, por lo tanto, no se registran en ella. Este tipo de operaciones internas suelen ser más rápidas y menos costosas, pero también menos seguras, ya que carecen del respaldo de la descentralización. Los *exchanges* centralizados emplean este sistema para mantener su contabilidad y garantizar la auditabilidad de sus registros. La información *off-chain* puede también referirse a cualquier otra información sobre las transacciones, como los precios de mercado a las que se han realizado, o la identidad del usuario que la ha realizado. En estos servicios centralizados la custodia de los fondos deja de estar en posesión del titular.

Por otro lado, los *Decentralized Exchanges* (*DEXes*) operan de manera descentralizada sin la intervención de una entidad central. Su funcionamiento se basa en *smart contracts* desplegados sobre la *blockchain*, lo que garantiza que todas las operaciones se ejecuten de forma automática, transparente y verificable. A diferencia de los CEX, los DEX no requieren la identificación de los usuarios mediante procesos KYC, y todas las transacciones realizadas en estas plataformas se registran *on-chain*, siendo accesibles públicamente y, por tanto, auditables por cualquier persona.

Tanto los CEX como los DEX ofrecen servicios de intercambio de criptoactivos (*swaps*) y el uso de puentes (bridges), que permiten la transferencia o conversión de activos digitales entre diferentes *blockchain*s (por ejemplo, de la red de Bitcoin a la red de Ethereum). Esta capacidad de interacción entre cadenas se conoce como interoperabilidad, y constituye un elemento clave en el desarrollo del ecosistema cripto. Sin embargo, solo los CEX permiten el intercambio de las criptomonedas por moneda fiduciaria, lo que se traduce en proveer de liquidez al usuario final.

Finalmente, el término Virtual Asset Service Providers (VASPs) o Crypto Asset Service Providers (CASPs) se emplea en el ámbito regulatorio, particularmente por organismos como el Grupo de Acción Financiera Internacional (GAFI) para designar a las entidades o personas jurídicas que ofrecen servicios relacionados con activos virtuales. Esto incluye la compra, venta, transferencia, almacenamiento o custodia de criptomonedas y criptoactivos, como es el caso de los *exchanges*.

Privacidad, seguridad y trazabilidad

Aunque las transacciones con criptomonedas son públicas y quedan registradas de manera permanente en la *blockchain*, los usuarios pueden mantener cierto grado de privacidad mediante diferentes estrategias. Entre las más comunes se encuentra el uso de múltiples billeteras para diferentes operaciones y la implementación de tecnologías como *coinjoins* y *mixers*, que permiten ofuscar el origen y destino de los fondos. Estas técnicas dificultan el seguimiento de los flujos de dinero, en el caso de las criptomonedas, o del 'valor', en el caso de los tokens y otros criptoactivos, los cuales, aunque no representen necesariamente una moneda, poseen un precio de mercado y pueden ser intercambiados o vendidos en diversas operaciones. En la *blockchain*, los usuarios no se identifican por su nombre real sino mediante direcciones alfanuméricas que actúan como identificadores únicos y no están asociadas de manera directa a la identidad personal del usuario. Por tanto, se dice que las operaciones en las redes de criptomonedas son pseudónimas, las entidades participantes "se ocultan" bajo una o varias direcciones sin revelar su identidad real.

Sin embargo, la privacidad que ofrece este sistema no es absoluta. Existen herramientas de analítica de *blockchain* que pueden analizar las transacciones e identificar agrupamientos o "clúster" de direcciones. Un clúster se define como un conjunto de direcciones que, a partir de diversos indicadores heurísticos o analíticos, se considera que podrían pertenecer a una misma persona o entidad. Esta identificación puede realizarse de forma automatizada, mediante algoritmos incorporados en las herramientas de análisis, o bien a partir de la interpretación y conocimiento contextual de analistas e investigadores especializados.

TÁCTICAS, TÉCNICAS Y PROCEDIMIENTO INVOLUCRADOS

La identificación de TTP constituye un componente fundamental en el estudio de las campañas de desinformación, ya que permite comprender las dinámicas operativas que sustentan su planificación y ejecución. La aplicación de marcos metodológicos como DISARM ofrece una estructura analítica que facilita la descomposición de los componentes de estas campañas, la detección de patrones de comportamiento y la identificación de relaciones entre actores, medios y recursos. Este enfoque no solo busca reconocer las TTP empleadas, sino también desarrollar estrategias de mitigación destinadas a reducir su efectividad y prevenir su reproducción. En este sentido, el análisis de TTP se consolida como una herramienta esencial para transformar la información obtenida en inteligencia accionable, orientada al diseño de procedimientos y acciones frente a la desinformación.

Identificación de las TTP en función de la clasificación de actores

El análisis de las TTP, junto con los distintos modus operandi empleados en las campañas de desinformación, permite comprender como distintos tipos de actores configuran y ejecutan sus operaciones en el entorno digital. La clasificación de estos actores, que abarca desde grupos APT, hasta colectivos hacktivistas y actores de cibercrimen, posibilita identificar patrones de actuación en función de sus motivaciones, capacidades y objetivos estratégicos.

En el **ANEXO I** se describen las TTP identificadas a nivel general entre las diferentes tipologías de actores, mapeadas de acuerdo con el marco metodológico DISARM, lo que posibilita vincular las TTP observadas con las distintas fases operativas de las campañas de desinformación.

Amenazas Persistentes Avanzadas (APT)

En el seguimiento realizado entre los diferentes actores desinformativos con intereses en España, se ha considerado identificar a cuatro de ellos por su mayor relevancia y capacidad para desarrollar e impulsar campañas de desinformación/FIMI en España.

Estos actores son China, Rusia e Irán, cada uno de ellos con una motivación diferente (*T0074 - Determine strategic ends*), en línea con los patrones descritos en European External Action Service (2022), razón que influye sobremanera en la elección del público

objetivo. Todos ellos tienen en común la instrumentalización de la diáspora, así como los sucesos mediáticos que suceden en España y que les permite crear narrativas que favorezcan sus intereses (*T0068 - Respond to breaking news event or active crisis*), tal como se observa en European External Action Service (2023). La red Pravda que publicó casi 300 contenidos en cinco días sobre los disturbios anti-inmigrantes en Torre Pacheco. Asimismo, diversas investigaciones confirman también la implicación de Irán en operaciones de influencia y desinformación (Yasur y Citrinowicz, 2024; Wright, 2025; Trilateral Research, 2025), cuyos comportamientos pueden interpretarse, desde un punto de vista analítico, como alineados con las dinámicas propias de las técnicas T0074 y T0068 anteriormente mencionadas, aun cuando estas fuentes no utilicen explícitamente dicha taxonomía.

Entre las principales finalidades que persiguen las FIMI de China, Rusia e Irán se encuentra el aumento de su prestigio internacional (*T0136.008 - Increase Prestige*), coincidente con los objetivos estratégicos identificados en European External Action Service (2025), en este caso, dentro de la sociedad española. Uno de sus objetivos fundamentales es contrarrestar la hegemonía geopolítica de Estados Unidos para, según sus argumentos, cambiar a un modelo multipolar (European External Action Service, 2025). En este contexto, Irán comparte con los dos primeros su interés en debilitar la influencia occidental, tal como evidencian sus operaciones de influencia documentadas en Yasur y Citrinowicz (2024), aunque se diferencia por su dimensión ideológica y religiosa, a través de la cual busca proyectarse como una potencia legítima y víctima de la agresión extranjera. Sus campañas suelen apoyarse en la producción de contenidos adaptados culturalmente (*T0101 - Create Localised Content*), una dinámica ampliamente observada en los estudios sobre FIMI recogidos en Wright (2025) y Trilateral Research (2025), y en la creación de identidades falsas y redes automatizadas (*T0097 - Expert Persona; T0146 - Account Asset*), elementos recurrentes en las operaciones iraníes analizadas por Yasur & Citrinowicz (2024), combinando la desinformación con acciones de presión internacional.

China

Una de las técnicas más comunes empleadas es la de cultivar apoyo (*T0136 - Cultivate Support*), sirviendo de fundamento para posicionarse como un modelo económico y social reconocido globalmente y así lograr que su influencia se expanda desde el Indo-Pacifico al resto del Mundo (Bomassi, 2025). China vigila la posición de la Unión Europea con respecto a la guerra en Ucrania para no ver mermadas sus relaciones económicas con el bloque comunitario.

En esta dirección surgen numerosas narrativas que se pueden ver identificadas en los datos mostrados y que sugieren posibles divisiones en el seno de la Unión Europea respecto a su posición con Estados Unidos (política arancelaria de Trump):

- Pretenden ofrecer una visión de rivalidad y competencia entre los países de la Unión Europea –una dinámica alineada con las técnicas T0066 (Degrade Adversary) y T0079 (Divide)–, tal como reflejan distintas narrativas en medios chinos (Zheng, 2025; Guancha, 2025a; People's Daily, 2025a).

- Referente al conflicto en Ucrania, enfatizan ciertas contraposiciones respecto a la actuación de la OTAN, como el gasto militar en España (Yang, 2025a; Shea, 2025).

China canaliza su poder a través de la cultura, fomentando actividades culturales y sociales en el plano físico (promoción de la celebración del año nuevo chino en España) y de las relaciones económicas, potenciando la presencia de empresas en el mercado chino, ya sea a través del comercio, las ventas o la inversión (International Republican Institute, 2023), con ejemplos visibles en iniciativas culturales promovidas en España (Spain-China Foundation, 2025).

A nivel comercial hacen especial énfasis en las buenas relaciones que siempre han existido entre España y China. El sector automovilístico chino de vehículos eléctricos e híbridos se impone en España a través de guerra de precios y con agresivas campañas de publicidad, con reportajes ensalzando las virtudes de ese mercado con respecto al resto, para posicionarse en ventaja geopolítica (Banco de España, 2024), como evidencian diversas narrativas mediáticas que refuerzan la superioridad del vehículo eléctrico chino (Guancha, 2025b; People's Daily, 2025b).

Por contra; se observa la utilización de narrativas para menoscabar la confianza de los ciudadanos españoles en las siguientes áreas:

- Resaltar la vulnerabilidad en sectores tan críticos como el sistema energético español. A la vez que ensalzan las virtudes de sus avances tecnológicos en

materia energética (tecnología solar), tal como recogen publicaciones sobre tecnología solar (Yang, 2025b).

- Ofrecen una visión de nuestro país como títere de Estados Unidos, que sucumbe a la presión ejercida por el Gobierno de ese país y de la Unión Europea con respecto decisiones soberanas de nuestro país –caso Huawei–, una narrativa alineada con la técnica T0066 (Degrade Adversary) respaldada por análisis previos (Esteban y Armanini, 2023; Márquez de la Rubia, 2025) y reforzada por abundante propaganda en medios afines al Partido Comunista Chino (Guancha, 2025c; Ifeng News, 2025; Southcn News, 2025; People's Daily, 2025c; Hoeren, 2025).

Rusia

El país euroasiático mantiene una narrativa constante centrada en acusar a la sociedad española de racista (Szakács y Bognár, 2021), con el propósito de presentar la inmigración como una amenaza entre los propios españoles, proyectar la imagen de una Europa insegura para los extranjeros y poner de manifiesto las carencias de las políticas migratorias de la Unión Europea. Un ejemplo de esta estrategia se evidenció en los incidentes ocurridos en Torre Pacheco (Maldita, 2025), los cuales fueron aprovechados para reforzar esta línea de desinformación, enmarcada en las técnicas *Cause Harm (T0140) y Undermine (T0135*) señaladas por el European External Action Service (2023). Además, recurre a los siguientes procedimientos:

- Creación de diarios digitales con plataformas en diferentes lenguas oficiales españolas mostrando incluso diferentes noticias locales para aumentar su difusión en España (*T0152 - Digital Content Hosting Asset*) según el European External Action Service (2024b), mediante la proliferación de sitios web como *Catalan News Pravda*, *Basque News Pravda* y *Galician News Pravda*, que replican contenidos alineados con la narrativa rusa adaptándolos a idiomas autonómicos para ampliar su alcance.

- Instrumentalización de la misma narrativa observada en China con respecto a las diferencias de opinión hacia el porcentaje de gasto militar comprometido con la OTAN con especial énfasis en el caso español (European External Action Service, 2024a), mediante coberturas en medios alineados con Rusia como TopWar (2025a), RT (2025a) y ESRT (2025), presentando a una España dividida y reticente a asumir compromisos de defensa.

- Exponen la imagen de una España caótica y de desconfianza en las Instituciones, mediante la difusión de contenido, mostrando protestas públicas relacionadas con las dificultades de acceso a la vivienda de los ciudadanos,

oposición a la amnistía sobre el independentismo y otros problemas sociales relacionados (Govorit Moskva, 2025; ZakonVremeni, 2025; Pashin, 2025).

- Difusión de mensajes dirigidos a acentuar el desacuerdo respecto a la posición que toman la Unión Europea y Estados Unidos frente a la guerra en Ucrania, una acción que se ajusta a la técnica *Divide* (*T0079*), según Aukia y Kubica (2023).

- Instrumentalización de eventos nacionales como los ocurridos en el depósito de la fábrica Rheinmetal en Murcia, presentándolos como acto de sabotaje con el objetivo de debilitar el apoyo logístico de España a Ucrania y distorsionar los hechos (T0076 - Distort), tal como difundieron diversos medios digitales alineados con Rusia como por ejemplo News-Kiev (2025) y Spanish News Pravda (2025).

- Manipulación informativa sobre el apagón generalizado ocurrido en España en abril de 2025, mediante la difusión de diversas narrativas. Entre otras emplearon el relato de "falta energía porque no le compramos gas a Rusia"; intentado posicionarse como fundamental en el suministro energético de Europa reduciendo así su independencia energética (RT, 2025b; TopWar, 2025b; YaOstrov, 2025).

- Igualmente, como al principio del apagón se especuló con la posibilidad de un ataque ruso, al demostrase que no había sido tal, se usó la TTP (*T0075.001 - Discredit Credible Sources*) con el objetivo de desacreditar medios fiables que había publicado dicha posibilidad, además de victimizarse (European External Action Service, 2023).

Irán

Irán se configura como un actor estatal que utiliza de forma coordinada la desinformación, la ciberinteligencia y las operaciones de influencia informativa con el objetivo de reforzar su posición geopolítica y debilitar las narrativas occidentales, una dinámica coherente con la técnica T0074 - Determine Strategic Ends (Microsoft, 2023b, 2024a; CyberProof, 2025).

Los grupos APT vinculados al país destacan por la creación de identidades falsas y la suplantación de expertos o instituciones, siguiendo patrones como *T0097.108 - Expert Persona* y *T0097.204 - Think Tank Persona* (Gatewatcher, 2025; Lakshmanan, 2025). Estas actividades se combinan con campañas de phishing y con la producción de contenidos adaptados culturalmente, en línea con *T0101 - Create Localised Content* (Microsoft, 2023b;

Gatewatcher, 2025). Tales acciones buscan influir en debates políticos, especialmente en torno a la guerra en Gaza o a la respuesta internacional tras los ataques israelíes de 2025, conforme a la técnica T0068 - *Respond to Breaking News Event or Active Crisis* (CyberProof, 2025; Global Influence Operations, 2025).

Su modus operandi se apoya en redes de bots y portales que amplifican mensajes alineados con los intereses de Teherán, asociados a *T0146 - Account Asset y T0098 - Establish Inauthentic News Sites* (Sha, 2025; Global Influence Operations, 2025). Estas operaciones contribuyen a polarizar y a desacreditar fuentes de información fiables, en consonancia con *T0079 – Divide y T0075.001 - Discredit Credible Sources*, integrando la manipulación de crisis con narrativas que presentan a Irán como un actor estabilizador y víctima de la agresión occidental (CyberProof, 2025; Global Influence Operations, 2025; Sha, 2025).

En el contexto europeo, y de manera indirecta en España, se ha observado el uso de comunidades afines o disidentes como vectores de influencia, de acuerdo con *T0010 - Cultivate Ignorant Agents* (England, 2025), junto con actividades de intimidación hacia críticos del régimen, relacionadas con *T0140.002 - Intimidate* (Reporters Without Borders, 2025).

Asimismo, la ocultación de activos y las operaciones mediante proxies, estructuras cifradas y estrategias de negación de responsabilidad permite a Irán mantener una presencia informativa persistente y de difícil atribución, responden a técnicas como *T0128 - Conceal Information Assets* y *T0129 - Conceal Operational Activity* (Newman, 2024; Microsoft, 2024b).

Todo ello consolida su papel como un actor híbrido con creciente capacidad para articular presión política, operaciones digitales y manipulación narrativa en el marco de su estrategia exterior (Microsoft, 2023b; CyberProof, 2025).

Hacktivismo

Los grupos hacktivistas desarrollan campañas de desinformación que combinan ciberataques técnicos, como el *defacement*, los *ataques DDoS* o la *filtración de información*, con sofisticadas operaciones de manipulación informativa y sociopolítica. El objetivo principal es maximizar el impacto reputacional y emocional en la sociedad, erosionando la confianza pública en instituciones, gobiernos y medios de comunicación.

Para ello, estructuran sus acciones en varias fases, comenzando por la planificación estratégica, donde definen si buscan ventajas geopolíticas, políticas internas o ideológicas, y seleccionan cuidadosamente a sus víctimas en función de su valor mediático. Por ejemplo, durante elecciones al Parlamento Europeo, han atacado webs de partidos políticos y organismos electorales para generar un impacto inmediato y debilitar la confianza en el proceso democrático.

En la fase de diseño y desarrollo de la campaña, los hacktivistas aprovechan cámaras de eco y burbujas de filtro en comunidades online para reforzar sus mensajes y captar nuevos simpatizantes. Generan y difunden contenido textual, visual y audiovisual, como mensajes reivindicativos, memes, imágenes y vídeos manipulados mediante inteligencia artificial (*deepfakes* y *cheap fakes*), distorsionando hechos y desacreditando fuentes oficiales.

Un ejemplo de esto es la difusión de memes que ridiculizan a líderes europeos presentándolos como marionetas de Estados Unidos, o la creación de vídeos generados por IA en los que políticos aparecen en situaciones absurdas o diciendo lo contrario a lo que defienden públicamente. Además, han coordinado ataques para desacreditar agencias de *fact-checking*, acusándolas de manipulación durante crisis informativas, y han explotado temas divisivos, como el debate sobre el aborto, lanzando mensajes simultáneos a favor y en contra para dificultar puntos de encuentro y polarizar a la sociedad. La movilización y financiación de estas campañas se apoya en la captación de miembros y simpatizantes, invitando a la participación activa en ataques o en la difusión de información relevante para el grupo. Utilizan plataformas digitales y criptomonedas para recaudar fondos, mezclando motivaciones políticas y económicas. Por ejemplo, en canales vinculados a campañas sobre Palestina, se han difundido direcciones de criptomonedas solicitando donaciones "para sostener la resistencia digital", mostrando cómo el *hacktivismo* puede mezclar la motivación política con la necesidad de recursos económicos.

Durante la ejecución y amplificación de la campaña, los hacktivistas publican contenido en activos propios y canales de difusión unidireccionales, empleando memes y anuncios para ridiculizar y desacreditar a sus objetivos. También promocionan y coordinan eventos

físicos y acciones simbólicas para reforzar sus narrativas y movilizar a la audiencia. En el contexto de conflictos internacionales, como el de Gaza, han convocado protestas sociales y campañas de odio étnico o religioso, polarizando a la población y aumentando la visibilidad del conflicto.

Finalmente, en el ámbito de las operaciones ofensivas y el control del entorno informativo, los grupos hacktivistas llevan a cabo acciones técnicas como phishing, DDoS o intrusiones para manipular la información visible, suprimir críticas y dominar el discurso en línea. Han atacado sitios de verificación o periodistas críticos, sobrecargando sus sistemas o filtrando credenciales para impedir la publicación de desmentidos y asegurar que su narrativa sea la dominante en el espacio informativo. Tras analizar en detalle las motivaciones, el modus operandi y las distintas formas de actuación de los grupos hacktivistas en el ámbito de la desinformación, resulta fundamental comprender cómo se articulan sus acciones a lo largo de todo el ciclo de una campaña.

Cibercrimen

Los grupos de cibercrimen que utilizan la desinformación como herramienta clave estructuran sus campañas siguiendo una lógica basada en fases, comenzando por la planificación estratégica. En esta etapa, definen objetivos como la obtención de beneficios económicos a través de la extorsión y el fraude, o la manipulación política para perpetuar comunidades y ecosistemas digitales afines. Un ejemplo ilustrativo es la proliferación de campañas de inversión fraudulenta en entornos como *Web3*, *blockchain* o la *Dark Web*, donde se promueven productos inexistentes o se facilita la piratería de contenido digital, generando confianza en la audiencia mediante el uso del sesgo de autoridad o la ilusión de conocimiento.

La segmentación de la audiencia es un proceso sofisticado, donde los cibercriminales identifican tanto a usuarios generalistas, vulnerables por su desconocimiento tecnológico, como a usuarios especializados, a quienes manipulan reforzando su exceso de confianza y validando sus creencias. Así, logran que ambos perfiles participen en esquemas de fraude, como ocurre en campañas de phishing dirigidas a usuarios poco familiarizados con nuevas tecnologías, o en estafas de inversión que explotan la confianza de quienes creen dominar el sector.

En la fase de preparación, los atacantes analizan y segmentan a la audiencia tanto demográfica como psicográficamente, localizando comunidades online susceptibles, como foros de criptomonedas o grupos de videojuegos. Aprovechan cámaras de eco y burbujas de filtro para reforzar sus mensajes y emplean técnicas de microsegmentación para adaptar el contenido a los intereses y vulnerabilidades de cada grupo. Por ejemplo, las

estafas vinculadas con NFT suelen dirigirse a usuarios jóvenes y activos en redes sociales, mientras que las campañas de phishing por correo electrónico tienden a orientarse hacia personas de mayor edad.

La creación de contenido es fundamental en estas campañas. Los cibercriminales desarrollan textos, imágenes y vídeos, a menudo generados por inteligencia artificial, para dotar de verosimilitud a sus narrativas. Recurren al uso de *deepfakes* y *cheap fakes* para manipular material audiovisual, así como a la reutilización o plagio de contenido preexistente para amplificar el alcance de sus mensajes. Un caso ilustrativo es la difusión de noticias falsas sobre oportunidades de inversión o tutoriales fraudulentos que inducen a la descarga de *malware*.

La selección de canales constituye igualmente una fase estratégica, donde los atacantes emplean desde redes sociales masivas como X (antes Twitter), Instagram o TikTok, hasta plataformas de mensajería instantánea como Telegram o Discord, pasando por foros especializados y servidores de videojuegos. Para reforzar la credibilidad de sus operaciones, establecen activos digitales como cuentas verificadas, *bots* automatizados y perfiles con imágenes generadas por IA o fotografías de personas atractivas, lo que facilita la captación y retención de víctimas.

En la fase de ejecución, los cibercriminales difunden su contenido de forma coordinada, utilizando redes de *bots* para amplificar mensajes y generar validaciones falsas mediante comentarios o reseñas. Mantienen su presencia en el entorno digital ocultando su infraestructura y las trazas de su actividad mediante pseudónimos, borrado de huellas digitales y el uso de criptomonedas para el blanqueo de capitales. Además, incentivan la compartición de sus mensajes a través de programas de afiliados o recompensas, logrando que las propias víctimas contribuyan de manera voluntaria a la expansión y legitimación de la campaña.

Estrategias de respuesta y mitigación sobre las TTP prioritarias

Tras la identificación de las TTP empleadas por los actores asociados a grupos APT, de *hacktivismo* y de cibercrimen, se desarrolló un análisis comparativo con el propósito de determinar aquellas con mayor recurrencia e impacto operativo. Este ejercicio de correlación permitió establecer una priorización de las TTP que, por su frecuencia e influencia, configuran un riesgo estratégico para la seguridad. El **ANEXO II** presenta una matriz comparativa que sintetiza los distintos TTP identificados en cada categoría de actor, así como su nivel de concurrencia.

Los resultados de este proceso evidenciaron que la técnica *T0068 - Respond to Breaking News Event or Active Crisis* es la más recurrente entre los actores analizados, al erigirse como un vector narrativo central en las campañas de desinformación que explotan los vacíos informativos durante crisis o eventos de alta exposición mediática. Por otro lado, se constató que las TTP asociadas a la generación de contenido sintético mediante inteligencia artificial, especialmente las vinculadas a *deepfakes*, configuran un riesgo emergente de primer orden, dada su creciente accesibilidad tecnológica y su capacidad para provocar daños reputacionales y políticos de gran alcance. En este contexto, se estudiaron las TTP asociadas a la creación de contenido específico mediante imágenes (TA06 - Develop Content) y vídeo (TA07 - Develop Video-Based Content). Ambas representan dominios operativos de alta prioridad, al concentrar las actividades más recurrentes orientadas a influir en la opinión pública a través de medios visuales. Los *cheap fakes*, elaborados con herramientas básicas, permiten una difusión rápida y masiva de mensajes alterados, mientras que los *deepfakes* introducen un grado de realismo avanzado gracias al uso de modelos de IA, incrementando su impacto emocional y dificultando su detección forense.

Como resultado, el presente trabajo selecciona dos TTP prioritarias para su estudio junto con la definición de medidas de mitigación específicas, siendo la T0068 - Respond to Breaking News Event or Active Crisis, por su reiteración y su papel amplificador en la desinformación, y la T0087.001 - Develop AI Generated Videos (Deepfakes) por su actual relevancia y potencial de expansión como amenaza. Ambas técnicas combinan la capacidad de generar material audiovisual sintético con la oportunidad de difundirlo en momentos de máxima sensibilidad mediática, generando un escenario de riesgo que exige una respuesta coordinada y proactiva.

Para ello, se propone la creación de una estrategia organizativa con competencias en verificación forense, trazabilidad de contenidos mediante estándares como C2PA (Coalition for Content Provenance and Authenticity, 2023), monitorización anticipada de campañas maliciosas y coordinación directa con plataformas de gran escala (VLOPs y VLOSEs). Esta iniciativa, en consonancia con la DSA (Regulation (EU) 2022/2065,

2022) y el Reglamento de Inteligencia Artificial –AI Act– (Regulation (EU) 2024/1689, 2024), permitiría consolidar una infraestructura nacional capaz de detectar y mitigar eficientemente las amenazas derivadas de la manipulación audiovisual avanzada, fortaleciendo la resiliencia frente a la desinformación más técnica.

Análisis de la técnica T0068 - Respond to Breaking News Event or Active Crisis

La técnica *T0068 - Respond to Breaking News Event or Active Crisis*, aborda la explotación intencional de situaciones de crisis o eventos de alta atención pública como vehículos para difundir contenido manipulado. Los actores que emplean esta TTP aprovechan el vacío informativo inicial, cuando las fuentes oficiales aún no se han pronunciado y la sociedad busca respuestas inmediatas, para introducir falsas narrativas que modelan la interpretación colectiva de los hechos.

La eficacia de esta técnica se basa en la sincronización y automatización, donde los atacantes preparan de antemano material alterado, como por ejemplo vídeos, declaraciones o imágenes manipuladas, que publican de forma casi simultánea al suceso real, aprovechando el alto nivel de atención mediática. El contenido se acompaña de un lenguaje emocionalmente cargado y de elementos visuales que imitan la identidad gráfica de instituciones oficiales o medios de comunicación, con el propósito de dotarlo de credibilidad. La difusión coordinada mediante cuentas falsas genera picos anómalos de actividad detectables mediante análisis temporal y modelado de propagación digital (Li et al., 2020; Boháček y Farid, 2024).

La detección y mitigación de esta técnica exige sistemas de vigilancia multimodal en tiempo real, capaces de integrar el análisis de texto, imagen y audio, junto con protocolos institucionales de respuesta rápida. Una vez identificado el contenido sospechoso, se evalúa su nivel de riesgo y se activa la cooperación con plataformas digitales bajo el marco del DSA y el AI Act. Asimismo, la vigilancia post-crisis resulta esencial para impedir la reaparición o mutación del contenido desinformativo. Mediante técnicas de *hashing perceptual y análisis semántico*, complementadas con herramientas de trazabilidad digital como C2PA y Truepic, se garantiza la detección de variantes manipuladas o recon-textualizadas. Estas medidas se enmarcan en las estrategias de "content provenance" impulsadas por la Comisión Europea.

La técnica T0068 pone así de manifiesto que la lucha contra la desinformación no depende solo de la verificación posterior, sino de la capacidad de anticipación, coordinación y respuesta inmediata de las instituciones frente al uso malicioso de la información en contextos de crisis. En el **ANEXO III** se presenta en detalle el análisis sobre la detección y mitigación de las TTP *T0068: Respond to Breaking News Event or Active Crisis.*

Análisis de la técnica T0087.001 - Develop AI Generated Videos (Deepfakes)

La técnica *T0087.001 - Develop AI Generated Videos (Deepfakes)* describe el uso de modelos de inteligencia artificial para crear o modificar vídeos en los que se simulan rostros, voces o movimientos de personas reales con una apariencia auténtica. A diferencia de las imágenes estáticas, los vídeos *deepfake* integran continuidad temporal, expresión facial y sincronía de audio, lo que incrementa su realismo y su capacidad de manipulación psicológica. Su utilización con fines maliciosos constituye uno de los desafíos más significativos para la seguridad en la actualidad.

Esta TTP combina diversos procesos tecnológicos, como el *face swapping* (intercambio facial), el *lip-sync reenactment* (sincronización labial con audio clonado) y el *motion transfer* (transferencia de gestos y movimientos). Estas técnicas emplean redes neuronales que aprenden la estructura facial y gestual de un individuo a partir de material de entrenamiento, generando posteriormente secuencias manipuladas que imitan sus patrones de habla o comportamiento. El resultado es un vídeo que aparenta autenticidad incluso ante observadores expertos, lo que dificulta la detección tanto humana como automatizada del caso.

Desde una perspectiva operativa, los *deepfakes* en vídeo se utilizan para manipular declaraciones atribuidas a autoridades, difundir falsos comunicados institucionales o influir en procesos electorales. Los atacantes buscan alterar la credibilidad pública en momentos clave, maximizando el impacto reputacional y emocional. La velocidad de propagación en redes sociales amplifica su efecto, ya que un vídeo falso difundido durante las primeras horas de un evento crítico puede condicionar la opinión pública antes de que existan desmentidos oficiales.

Las señales de detección de esta técnica combinan análisis visual, auditivo y temporal. A nivel visual, pueden detectarse incongruencias en la orientación de la cabeza (*head pose*), irregularidades en la sincronización labial y artefactos de mezcla o *blending* en los bordes faciales. A nivel auditivo, la desincronización entre fonemas y movimientos labiales constituye un indicador temprano de manipulación, identificable mediante el contraste entre transcripciones generadas por reconocimiento de voz (*Automatic Speech Recognition*) y lectura labial automatizada (*lip Reading*).

Finalmente, los modelos de análisis temporal, como los *Vision Temporal Transformers* o *detectores multimodales A/V*, son capaces de detectar patrones dinámicos anómalos en la secuencia audiovisual que escapan al ojo humano.

Las estrategias de mitigación se estructuran en tres ejes complementarios. En primer lugar, el eje tecnológico, que requiere el desarrollo de detectores multimodales capaces de identificar inconsistencias entre audio y vídeo, así como la integración de estándares de trazabilidad como C2PA y Content Credentials en los vídeos oficiales. En segundo lugar, el eje operativo, centrado en la cooperación con plataformas digitales para implementar protocolos de retirada rápida (*takedown*) y etiquetado obligatorio de contenido generado por IA, conforme a lo dispuesto en el *AI Act*. Por último, el eje institucional y preventivo, que abarca la creación de repositorios de muestras verificadas, el reentrenamiento continuo de modelos de detección y la aplicación de medidas de *pre-bunking* comunicacional para anticipar narrativas manipuladas.

La relevancia de esta técnica no reside únicamente en su potencial para falsificar la realidad, sino en su capacidad para erosionar el valor probatorio de las evidencias audiovisuales. En un entorno donde cualquier vídeo puede ser cuestionado como posible deepfake, la frontera entre la verdad y la falsificación se difumina, afectando la confianza en los medios y las instituciones. En consecuencia, la gestión de esta TTP exige una respuesta que combine la innovación forense, la cooperación internacional y una política pública de transparencia digital sostenida. En el **ANEXO IV** se presenta en detalle el análisis sobre la detección y mitigación de las TTP *T0087.001 - Develop AI Generated Videos (Deepfakes)*.

PRODUCCIÓN DE INTELIGENCIA OPERATIVA SOBRE DESINFORMACIÓN

La producción de inteligencia operativa en el ámbito de la desinformación tiene como objetivo generar conocimiento verificable y aplicable sobre actividades hostiles que utilicen medios digitales para manipular o distorsionar la información. Desde una perspectiva técnica, este proceso se basa en la recopilación, correlación y análisis de datos procedentes de múltiples fuentes, con el objetivo de identificar patrones de comportamiento, infraestructuras asociadas y vínculos entre actores.

El panorama actual evidencia una creciente convergencia entre las operaciones en el ciberespacio y las campañas de desinformación, en la que grupos APT, hacktivistas y actores del cibercrimen utilizan la información como un vector operativo dentro de estrategias híbridas. Estas actividades combinan componentes técnicos y comunicativos orientados a influir en la percepción pública, debilitar la confianza institucional y promover intereses estratégicos determinados.

En este contexto, la inteligencia operativa permite identificar y contextualizar la amenaza desde una perspectiva técnica, incorporando además la evaluación de la intencionalidad y el impacto para generar productos orientados a la detección temprana y mitigación de riesgos. El resultado es una inteligencia estructurada, capaz de transformar datos fragmentados en conocimiento accionable, fortaleciendo la anticipación y la resiliencia frente a las campañas de desinformación coordinadas, con independencia de la naturaleza o la motivación del actor que las ejecute.

Clasificación y normalización de la información

El proceso de clasificación y normalización de la información constituye un elemento esencial en la producción de inteligencia operativa, ya que permite transformar datos dispersos en conocimiento estructurado y verificable. Su propósito es garantizar la coherencia, trazabilidad y utilidad práctica de la información, asegurando su correcta explotación analítica e interoperabilidad entre organismos y plataformas. La clasificación organiza los datos según su naturaleza y nivel de aplicación dentro del ciclo de inteligencia. En el contexto de la desinformación, esta puede estructurarse en tres niveles complementarios:

- **Nivel estratégico:** orientado a la identificación de los objetivos, motivaciones y alineamientos geopolíticos de los actores, así como al análisis de las narrativas

de largo alcance que buscan alterar percepciones o generar influencia sostenida.

- **Nivel operacional:** centrado en la caracterización de campañas, redes de difusión, plataformas empleadas y mecanismos de amplificación y coordinación entre canales.

- **Nivel táctico:** enfocado en los elementos técnicos específicos, como dominios, direcciones IP, certificados, cuentas automatizadas o indicadores de compromiso (IoC), que conforman la base observable de las operaciones, así como en la identificación de y vinculación de los actores responsables de los ataques desinformativos.

A su vez, la clasificación debe adaptarse a la tipología del actor, ya que cada perfil presenta técnicas dinámicas y operativas totalmente diferenciadas:

- **Grupos APT.** La información relativa a estos actores se categoriza en función de la infraestructura técnica empleada —como dominios, servidores de comando y control (C2), certificados digitales o herramientas—, así como por los patrones operativos que caracterizan su actividad, entre ellos la persistencia, la sincronización de campañas y la coincidencia de indicadores de compromiso. Su análisis también incorpora la identificación de vínculos con intereses estatales o geopolíticos, dado que estos grupos suelen actuar bajo el patrocinio o la afinidad de gobiernos que buscan obtener ventajas estratégicas, políticas o militares. Se prioriza la trazabilidad técnica, la atribución contextual y la integración de las operaciones de ciberespionaje, sabotaje y propaganda como componentes interrelacionados dentro de una misma estrategia de influencia.

- **Actores de cibercrimen.** La clasificación de este tipo de actores se orienta hacia el análisis de las estructuras de monetización que sustentan sus operaciones, incluyendo el uso de criptoactivos, foros underground y canales de transacción empleados para el intercambio de servicios ilícitos. Asimismo, se examina la infraestructura técnica reutilizada, como servicios de hosting o plataformas de anonimización, que facilitan la continuidad y el encubrimiento de sus actividades. En este contexto, adquieren especial relevancia las campañas de fraude, extorsión y manipulación reputacional asociadas a desinformación.

- **Hacktivismo.** La clasificación de los grupos hacktivistas se centra en la coordinación entre los componentes narrativos, simbólicos y técnicos de sus acciones. Estos actores articulan su actividad a través de canales públicos como X (anteriormente Twitter), Facebook o Telegram, donde difunden mensajes

reivindicativos y coordinan la amplificación de sus campañas. Los principales indicadores incluyen la sincronización entre operaciones técnicas (como ataques DDoS o *defacements*) y comunicativas, la cronología de sus intervenciones y los patrones de propagación de sus mensajes en redes sociales.

La normalización garantiza la integridad, coherencia y comparabilidad de la información clasificada dentro del proceso de producción de inteligencia. Este procedimiento se apoya en modelos estandarizados, como STIX, un lenguaje estructurado diseñado para representar la información sobre amenazas, que permite describir las relaciones entre objetos técnicos y contextuales, por ejemplo, la conexión entre un dominio, una red de bots y una narrativa determinada. Para facilitar un análisis transversal y una integración eficiente de los datos, la normalización debe basarse en una estructura común aplicable a todas las tipologías de actores, garantizando así su interoperabilidad y puedan correlacionarse en distintos niveles de análisis.

En términos operativos, esta estructura puede organizarse en los siguientes bloques:

- **Identificación del actor:** comprende la denominación, afiliación y motivación (política, económica, ideológica o híbrida), así como su nivel de organización, alcance geográfico y grado de vinculación con otros grupos. Este bloque resulta esencial para contextualizar la intencionalidad percibida de sus acciones dentro del ecosistema informativo.

- **Infraestructura y medios empleados**: incluye los elementos técnicos utilizados para la ejecución de las operaciones, tales como dominios, servidores, canales de comunicación, herramientas, certificados o plataformas de difusión. El análisis de esta infraestructura permite detectar patrones de reutilización, identificar dependencias tecnológicas y establecer conexiones entre distintas campañas o incidentes.

- **TTP utilizadas**: abarca las tácticas, técnicas y procedimientos desplegados, el grado de automatización aplicado y los vectores de ataque o influencia empleados. Este bloque facilita la comprensión del modus operandi de los actores, permitiendo relacionar sus capacidades técnicas con los objetivos estratégicos o comunicativos de la operación.

- **Narrativa dominante:** se centra en los temas recurrentes, los objetivos comunicativos, el idioma, el tono discursivo y la segmentación de la audiencia. El análisis de la narrativa permite identificar los marcos ideológicos y emocionales utilizados para orientar la percepción pública, así como las estrategias de legitimación o polarización adoptadas.

- **Contextualización de evidencias:** engloba los indicadores de compromiso, artefactos, metadatos y la cronología de actividades y eventos asociadas a amenazas y actores. Este bloque proporciona la base para la correlación de incidentes, la atribución técnica y la evaluación del impacto operativo de la campaña.

La aplicación de esta estructura común permite comparar actores heterogéneos bajo criterios unificados, detectar solapamientos operativos, identificar infraestructuras compartidas y establecer relaciones entre incidentes aparentemente aislados. De este modo, el proceso de clasificación y normalización no solo organiza la información, sino que transforma el conocimiento técnico en una base analítica sólida para la atribución, anticipación y mitigación de amenazas de desinformación, independientemente de la tipología o alcance del actor implicado.

Inteligencia accionable en MISP

La generación de inteligencia operativa frente a campañas de desinformación requiere un marco metodológico que permita transformar datos fragmentados en conocimiento estructurado, verificable y aplicable. En este sentido, MISP (MISP Project, s. f.) constituye una herramienta fundamental para la gestión integral de la información y la cooperación interinstitucional, al posibilitar la correlación de indicadores, la estandarización semántica mediante taxonomías y la integración de evidencias tanto técnicas como contextuales dentro de un entorno analítico común.

El uso de MISP en el ámbito de la desinformación amplía sus funciones tradicionales de ciberseguridad, al incorporar componentes narrativos y de influencia informativa junto a los puramente técnicos. Este enfoque integral permite representar en un único espacio la convergencia entre infraestructura digital, tácticas operativas y mensajes estratégicos, otorgando al analista una visión completa del ciclo de una campaña. De esta forma, MISP deja de ser un repositorio de indicadores para convertirse en un sistema de inteligencia accionable, donde la información técnica se combina con la interpretación contextual y la evaluación del impacto.

La plataforma se sustenta en un modelo de datos flexible que permite enlazar objetos complejos, como por ejemplo eventos, dominios, TTP, herramientas, narrativas o actores, describiendo sus relaciones de manera coherente y trazable. Este modelo se refuerza mediante taxonomías y esquemas de clasificación adaptables, que garantizan una coherencia analítica entre los niveles estratégico, operacional y táctico. Dichas taxonomías, ampliamente reconocidas por organismos como ENISA, Europol o la comunidad CSIRT,

posibilitan una interoperabilidad analítica entre diferentes instituciones y entornos de inteligencia.

Bajo este modelo, cada evento registrado en MISP puede estructurarse en cinco dimensiones esenciales de la operación, articuladas mediante taxonomías adaptadas al análisis de campañas de desinformación. Estas permiten estructurar de forma homogénea los componentes técnicos, comunicativos y contextuales, favoreciendo el análisis y la correlación entre diferentes fuentes:

- **Identificación del actor.** esta taxonomía abarca los atributos asociados con la tipología del actor (grupo APT, hacktivismo, cibercrimen), la afiliación, la motivación principal (política, económica, ideológica o híbrida) y el grado de persistencia operativa. A nivel analítico, puede incluir información relativa a la estructura del grupo, su historial de actividad, su grado de coordinación y la vinculación con entidades o ecosistemas de influencia. En el contexto de la desinformación, esta dimensión permite distinguir entre actores productores, amplificadores y facilitadores de contenido, ofreciendo una visión jerárquica y funcional de su papel dentro del ciclo de una campaña.

- **Infraestructura.** Esta taxonomía abarca tanto los elementos técnicos (dominios, direcciones IP, servidores, sistemas autónomos o servicios de hosting) como los medios de difusión y control de la información, incluyendo canales en redes sociales, plataformas de mensajería instantánea (Telegram, Discord, WhatsApp), blogs automatizados o portales webs. Asimismo, puede incorporar metadatos sobre distintos tipos de infraestructura, como redes de bots, proxys, content delivery networks (CDN) o certificados digitales reutilizados. En las campañas de desinformación, la identificación de estos activos permite trazar las rutas de propagación, detectar infraestructuras compartidas entre operaciones y establecer patrones de coordinación transnacional.

- **Tácticas, técnicas y procedimientos (TTP).** Referenciados mediante marcos estandarizados como DISARM o MITRE ATT&CK, esta taxonomía describe las fases operativas empleadas. En el contexto de la desinformación, las TTP pueden incluir desde la creación de contenido manipulado (T0023 - Distort Facts), la explotación de crisis o eventos de actualidad (T0068 - Respond to Breaking News Event or Active Crisis), hasta la amplificación automatizada de narrativas (T0119 - Cross-Posting). De igual modo, pueden documentarse técnicas de suplantación de identidad digital, enmascaramiento de fuentes o sincronización de publicaciones con operaciones cibernéticas paralelas, reflejando la naturaleza híbrida de este tipo de amenazas.

- **Tipología de narrativa.** Esta taxonomía recoge metadatos sobre el idioma, tono, temática y objetivo comunicativo del mensaje, así como sobre el tipo de audiencia a la que se dirige. Incluye, además, la categorización de narrativas dominantes, por ejemplo, antioccidental, antigubernamental, proconflicto, negacionista o victimista, y las estrategias discursivas utilizadas, tales como el humor, la emotividad o la autoridad pseudocientífica. A nivel técnico, permite asociar narrativas con canales o infraestructuras específicas, identificando clústeres temáticos o semánticos recurrentes. Esta dimensión es fundamental para analizar los mecanismos de polarización social, refuerzo ideológico y manipulación cognitiva que sustentan las campañas de influencia o desinformación.

- **Fiabilidad y credibilidad de evidencias.** Esta taxonomía evalúa la calidad, procedencia y coherencia técnica de los datos recopilados. Integra cronologías de eventos, junto con la evaluación de confianza de las fuentes, el grado de atribución y el nivel de correlación intercampaña. En el contexto de la desinformación, esta dimensión permite distinguir entre evidencias verificadas, parciales o especulativas, aplicando escalas de fiabilidad y credibilidad de las fuentes y su contenido, además de los sistemas de trazabilidad que garantizan la validación cruzada de la información. A nivel operativo, su adecuada normalización facilita la priorización de alertas, la depuración de falsos positivos y la elaboración de inteligencia de calidad basada en evidencia verificable.

La implementación de estos elementos mediante taxonomías y etiquetas personalizadas permite representar de forma unificada la dimensión técnica (infraestructura, TTP, vectores de ataque) y la dimensión narrativa (tema, idioma, audiencia, objetivo estratégico) de una operación. Un ejemplo de registro normalizado podría adoptar el siguiente formato:

disinfo:**actor**:actor-type=APT35

disinfo:**actor**:actor-affiliation=state_sponsored

disinfo:**actor**:actor-motivation=political_influence

disinfo:**actor**:actor-persistence=high

disinfo:**infrastructure**:platform=Telegram

disinfo:**infrastructure**:asset-type=communication_channel

disinfo:**infrastructure**:hosting-region=Europe

disinfo:**TTP**=T0068_respond_to_breaking_news_event_or_active_crisis

disinfo:**TTP**:automation-level=medium

disinfo:**narrative**:narrative-category=anti_NATO

disinfo:**narrative**:narrative-theme=geopolitical_legitimacy

disinfo:**narrative**:narrative-tone=persuasive

disinfo:**narrative**:narrative-audience=spanish_public

disinfo:**narrative**:narrative-language=es

disinfo:**evidence**:admiralty-code=A7

disinfo:**evidence**:source-platform=Telegram

disinfo:**evidence**:confidence-level=high

disinfo:**evidence**:correlation=linked_to_APT35_cluster

Este modelo demuestra cómo MISP permite documentar una campaña, combinando información técnica, contextual y narrativa, y facilitando la correlación de eventos con características comunes, como la reutilización de infraestructura, la coincidencia temática o la sincronización temporal entre operaciones, ofreciendo así una visión transversal de las actividades de influencia.

Además, MISP potencia la interoperabilidad analítica y el intercambio estructurado de información mediante su capacidad de exportar eventos y objetos en formatos estandarizados como STIX. Este mecanismo garantiza la compatibilidad semántica y la trazabilidad de los datos entre diferentes organismos o instancias federadas, permitiendo la correlación y el enriquecimiento mutuo de la información sin comprometer su integridad. En este marco, la inteligencia producida a través de MISP trasciende la función meramente descriptiva y adquiere valor operativo y estratégico. La correlación de datos técnicos con elementos narrativos posibilita la detección de convergencias entre campañas aparentemente independientes, la atribución de actividades a actores específicos y la anticipación de tendencias emergentes en el ámbito informativo.

En consecuencia, MISP se consolida como un sistema centralizado de análisis y cooperación interinstitucional, en el que la estandarización y la interoperabilidad de la información fortalecen la capacidad de detección, atribución y mitigación de amenazas de desinformación. La integración de datos técnicos, contextuales y narrativos bajo un mismo marco analítico genera inteligencia de alto valor operativo, refuerza la resiliencia institucional y optimiza la respuesta coordinada frente a campañas híbridas que combinan desinformación, ciberataques y manipulación mediática.

CONCIENCIACIÓN Y BUENAS PRÁCTICAS SOBRE DESINFORMACIÓN

La desinformación se ha convertido en un **r**iesgo estructural para la seguridad nacional y global, capaz de alterar percepciones, polarizar sociedades y erosionar la confianza en las instituciones (Comisión Europea, 2022a; DSN, 2021). En el marco de las operaciones híbridas, se integra como vector cognitivo dentro de estrategias de influencia que combinan ciberataques, manipulación mediática y el uso coordinado de campañas automatizadas a través de redes sociales.

Finalidad estratégica de la concienciación y la defensa cognitiva

La finalidad estratégica de la concienciación es reducir la superficie de ataque cognitiva fortaleciendo la alfabetización informativa, la capacidad crítica y los protocolos de verificación. La respuesta no puede limitarse al plano técnico ni basarse exclusivamente en indicadores de amenaza (TTP, MISP, STIX o similares), sino que requiere capacidades humanas sostenidas: educación, comunicación responsable y coordinación institucional. Asimismo, la desinformación y la manipulación informativa son el riesgo global más grave a corto plazo (2024–2026) y uno de los cinco principales a medio plazo. El informe advierte de la posibilidad de que la desinformación, potenciada por IA generativa, se convierta en el detonante de crisis institucionales o electorales. Por tanto, la concienciación se configura como un elemento esencial de la resiliencia nacional frente a amenazas híbridas y campañas de influencia (World Economic Forum, 2024).

Dimensión educativa y cognitiva

La educación y la alfabetización mediática son el núcleo de la resiliencia cognitiva frente a la manipulación informativa, constituyendo una capacidad esencial de la seguridad nacional. Organismos internacionales como la UNESCO (2021) y la Comisión Europea (2022a) coinciden en que la ciudadanía debe adquirir competencias críticas para identificar, contrastar y comunicar información de manera responsable. Fomentar estas habilidades contribuye no solo a fortalecer la cohesión social, sino también a consolidar una cultura democrática más resistente.

El desarrollo del pensamiento crítico, la identificación de sesgos cognitivos y el entrenamiento frente a la manipulación emocional constituyen objetivos prioritarios en la construcción de una ciudadanía informada ante la desinformación. En este sentido, el enfoque de *prebunking* de Linden (2020), basado en la exposición preventiva a ejemplos de desinformación y técnicas de manipulación, refuerza la capacidad de detección temprana y evita la propagación de contenidos falsos. Este enfoque, actualmente impulsado en programas europeos de alfabetización digital, se está incorporando también a iniciativas nacionales de ciberresiliencia y comunicación pública. Básicamente se centra en tres niveles de actuación, el cognitivo (reconocimiento de sesgos, emociones y patrones de manipulación), el técnico (aprendizaje del uso básico de herramientas de verificación mediante fuentes abiertas) y el conductual (fomento de hábitos responsables en la difusión de información).

Por lo que la educación mediática, integrada de manera transversal en sistemas formativos y programas institucionales, transforma la concienciación en una práctica sostenida, medible y verificable, situando al ciudadano como actor activo de la defensa nacional, convirtiéndose en una herramienta de defensa cognitiva colectiva y pilar del ecosistema de seguridad nacional.

Alfabetización digital y resiliencia cognitiva

Mientras que la alfabetización mediática se centra en el análisis crítico del contenido informativo, la alfabetización digital abarca el conocimiento técnico y operativo del ecosistema digital en el que dicho contenido se genera, manipula y distribuye. Su aplicación práctica implica enseñar a los usuarios a interpretar metadatos, identificar fuentes primarias y reconocer patrones de automatización en redes sociales y plataformas digitales (ENISA, 2024).

Esta formación debe incorporarse en tres ámbitos prioritarios: el educativo con una integración curricular en educación secundaria, universitaria y formación profesional, el administrativo con capacitación del personal público en comunicación segura y detección de bulos, y, el corporativo con programas internos de sensibilización sobre riesgos reputacionales y de ingeniería social. Estas capacidades refuerzan la resiliencia cognitiva nacional, tal y como se establece en la Estrategia de Seguridad Nacional y en los planes europeos de ciberalfabetización.

La educación mediática, digital y "figital" en entornos públicos y corporativos

Como extensión de la alfabetización digital, la educación mediática, digital y "figital" debe articularse como una política pública transversal, vinculada tanto a los planes de comunicación institucional como a los programas de ciberseguridad y cumplimiento normativo.

Las administraciones públicas y las empresas pueden actuar como multiplicadores de resiliencia, incorporando módulos formativos sobre verificación, trazabilidad y no amplificación en sus planes de concienciación. Estas acciones permiten reforzar la cohesión social, reducir la exposición a campañas de manipulación y preservar la confianza ciudadana en las fuentes legítimas.

En la práctica, esta integración puede materializarse a través de: (a) planes formativos para el personal público, que incluyan competencias de comunicación segura, verificación básica y detección de bulos, (b) programas de concienciación corporativa, orientados a la protección de la reputación digital, la gestión de riesgos comunicativos y la prevención de la ingeniería social, y finalmente, con (c) campañas institucionales de verificación ciudadana, diseñadas para fomentar el pensamiento crítico y la participación activa en la defensa cognitiva.

Por otro lado, la alfabetización "figital" añade una nueva dimensión a este enfoque, al integrar los comportamientos del mundo físico y digital dentro de un mismo marco de responsabilidad comunicativa. En este sentido, la conducta informativa de los individuos, tanto en línea como fuera de ella, se convierte en un elemento estratégico de seguridad colectiva.

Por lo que la concienciación, entendida como práctica continua y estructurada, convierte la información verificada en una defensa activa del espacio público y corporativo, alineada con las estrategias de seguridad nacional y europea. De este modo, las entidades públicas y privadas se consolidan como agentes activos de la defensa cognitiva nacional.

Verificación activa de contenido

Sin duda, la verificación constituye el núcleo operativo de la respuesta ante la desinformación. Basada en metodologías OSINT, documenta de forma trazable cada análisis e integra los resultados en plataformas de inteligencia como MISP (ENISA, 2024; DSN, 2021).

El objetivo de la verificación activa no se limita a determinar la autenticidad de un contenido, sino también a detectar patrones de repetición, amplificación y origen coordinado. Esta correlación entre el comportamiento digital y los indicadores técnicos posibilita identificar vínculos con campañas cognitivas u operaciones híbridas de influencia.

- Por ejemplo, una imagen viral puede verificarse mediante técnicas de búsqueda inversa y análisis detallado de metadatos, lo que permite determinar su publicación original, posibles manipulaciones gráficas, contexto temporal y medio de difusión inicial. Documentar estos pasos de forma trazable evita su difusión masiva y refuerza la respuesta institucional. Este proceso se complementa con un conjunto de herramientas especializadas de verificación y análisis digital, descritas en el **ANEXO V**, que abarcan desde la búsqueda inversa de imágenes hasta el análisis forense, reputación de dominios y detección de automatismos en redes sociales.

La verificación debe seguir un modelo estructurado, reproducible y conforme a los principios de trazabilidad, siguiendo el Esquema Nacional de Seguridad (ENS) y la Ley 40/2015, que exigen integridad, custodia y transparencia en los sistemas de información del sector público.

Procedimiento mínimo de verificación

La aplicación de un procedimiento garantiza un proceso verificable, transparente y jurídicamente sólido, permitiendo integrar los resultados en los sistemas de inteligencia técnica e institucional y fortaleciendo la capacidad nacional de respuesta ante campañas de manipulación informativa. Por este motivo es esencial seguir de manera sistemática el siguiente proceso:

- **Identificar y clasificar el contenido.** Determina el nivel de riesgo y la prioridad de análisis, permitiendo distinguir si se trata de una noticia, rumor, sátira, contenido manipulado o material generado por IA.

- **Evaluar la fuente y reputación del dominio.** Valorar la legitimidad y posible suplantación, analizando la autoría, registro, historial y nivel de fiabilidad.

- **Comprobar procedencia técnica.** Autenticar la evidencia, utilizando la búsqueda inversa, análisis de metadatos y archivos web que confirmen origen y cronología.

- **Contrastar el contexto.** Verificar coherencia narrativa y geográfica, revisando lugar, tiempo y concordancia con fuentes independientes.

- **Dictaminar y documentar.** Garantizar trazabilidad y custodia, clasificando el contenido como veraz, falso, engañoso o no verificable. Es esencial registrar y mantener una guía documentada con los hallazgos, herramientas empleadas, fecha, fuente y responsable.

Protocolos frente a narrativa sospechosa

Los protocolos frente a narrativas sospechosas definen las acciones de respuesta comunicativa que deben aplicarse cuando se detecta una campaña coordinada o un contenido potencialmente dañino. Su eficacia depende de tres factores: (a) rapidez, (b) coherencia institucional y (c) comunicación responsable (Comisión Europea, 2022b; DSN, 2021). El objetivo no es únicamente desmentir, sino neutralizar la capacidad de amplificación de la narrativa manipulada, preservando la confianza en las fuentes legítimas. Estos protocolos deben establecer flujos de actuación escalables desde el ciudadano hasta la administración pública o empresa afectada.

Fases operativas del protocolo

Cada actuación del protocolo debe conservar una trazabilidad completa, en cumplimiento de los principios de integridad, confidencialidad y conservación definidos por el Esquema Nacional de Seguridad (ENS), modelo que tomamos como referencia.

La verificación activa y la respuesta coordinada conforman así una doble capa de defensa cognitiva que permite anticipar, contener y desactivar operaciones de influencia antes de que alcancen impacto masivo. Por este motivo es imprescindible seguir las siguientes fases operativas:

- **Fase 1. Detección y verificación.** Identificar la narrativa sospechosa, registrar su origen, fuente y canal de difusión, y aplicar el procedimiento de verificación descrito anteriormente.

- **Fase 2. Evaluación y coordinación.** Analizar el alcance y el público afectado, valorar la intencionalidad del mensaje y coordinar la respuesta con las unidades competentes (gabinetes de comunicación, SOC, CSIRT o DSN).

- **Fase 3. Respuesta y documentación.** Emitir una comunicación institucional coherente, transparente y neutral, priorizando la claridad sobre la confrontación. Un ejemplo de comunicación puede verse en el **ANEXO VI.**

CONCLUSIONES

La desinformación representa hoy una amenaza que trasciende el ámbito comunicativo para convertirse en un instrumento de poder y desestabilización estratégica. Su capacidad para distorsionar percepciones, influir en la opinión pública y socavar la legitimidad de las instituciones la sitúa en el núcleo de las preocupaciones de seguridad en la actualidad. En un entorno donde las fronteras entre lo informativo, lo tecnológico y lo político se difuminan, las campañas de desinformación operan como mecanismos híbridos de influencia, combinando técnicas de ciberataque, manipulación narrativa y explotación algorítmica de plataformas digitales.

En este contexto, la respuesta institucional exige trascender los enfoques reactivos y avanzar hacia una gestión técnica y analítica de la información hostil, sustentada en una metodología estructurada de producción de inteligencia. Esta debe permitir no solo la identificación y trazabilidad de los contenidos manipulados, sino también el análisis relacional entre infraestructuras, actores implicados y narrativas propagadas. La aplicación de modelos basados en la correlación de datos, la detección de patrones operativos y la caracterización de TTP proporciona una base objetiva para transformar el volumen informativo en inteligencia accionable, orientada a la toma de decisiones estratégicas y la formulación de políticas de mitigación.

Plataformas como MISP resultan esenciales en este proceso, al ofrecer un marco técnico interoperable que posibilita la agregación, normalización y correlación automatizada de indicadores técnicos y contextuales. Su integración con marcos metodológicos como DISARM o MITRE ATT&CK permite describir de forma estandarizada los procedimientos empleados por actores de desinformación, facilitando la atribución, el análisis comparativo y la cooperación interinstitucional. Este enfoque impulsa una visión multidimensional de la amenaza, en la que la infraestructura técnica, la narrativa y la intencionalidad estratégica se analizan de manera conjunta, reforzando la capacidad de anticipación y respuesta coordinada.

Paralelamente, la dimensión cognitiva de la resiliencia social adquiere una relevancia creciente. La educación mediática y digital constituye una herramienta de prevención estructural que fomenta el pensamiento crítico, la autonomía informativa y la resistencia psicológica ante la manipulación. Estrategias como el *prebunking* y la integración de la alfabetización informacional en los sistemas educativos y comunicativos nacionales convierten la concienciación en una práctica esencial. Al mismo tiempo, dotar a la ciudadanía de competencias para analizar y verificar fuentes contribuye a construir un ecosistema comunicativo más seguro y responsable.

En definitiva, la lucha contra la desinformación requiere un enfoque integral y multidisciplinar, que combine el rigor técnico de la inteligencia operativa con la dimensión ética y formativa de la alfabetización mediática. La creación de metodologías técnicas normalizadas, la cooperación entre organismos y la educación crítica de la población conforman los pilares de una defensa cognitiva nacional eficaz y sostenible. Solo mediante la articulación de estos elementos será posible convertir la información en conocimiento verificable, la inteligencia en acción estratégica y la sociedad en un sujeto resiliente frente a la manipulación y la injerencia informativa.

REFERENCIAS BIBLIOGRÁFICAS

Adler, S., Hitzig, Z., Jain, S., Brewer, C., Srivastava, V., Christian, B., & Trask, A. (2024). Personhood credentials: Artificial intelligence and the value of privacy-preserving tools to distinguish who is real online. https://doi.org/10.48550/arXiv.2408.07892

Aldama, Z. (2020). Manual chino para después de la pandemia. Revista 5W. https://www.revista5w.com/temas/salud/manual-chino-para-despues-de-la-pandemia-8147

Annabell, T., Bishop, S., & Goanta, C. (2025). "You and TikTok are, and will remain at all times, independent contractors". Internet Policy Review, 14(3). https://doi.org/10.14763/2025.3.2014

Aukia, J., & Kubica, L. (2023). Hybrid CoE Research Report 8: Russia and China as hybrid threat actors: The shared self-other dynamics, pp. 25–28. European Centre of Excellence for Countering Hybrid Threats. https://www.hybridcoe.fi/wp-content/uploads/2023/04/NEW_Hybrid_CoE_Research_Report_8_web.pdf

Badillo, Á., Arteaga, Félix (2024). El impacto estratégico de la desinformación en España. Real Instituto Elcano. https://www.realinstitutoelcano.org/informes/informe-iberifier-el-impacto-estrategico-de-la-desinformacion-en-espana/

Banco de España. (2024). Boletín Económico 2024/T4: Proyecciones macroeconómicas e informe trimestral de la economía Española. https://www.bde.es/f/webbe/SES/Secciones/Publicaciones/InformesBoletinesRevistas/BoletinEconomico/24/T4/Fich/be2404-art03.pdf

Benedicto, M. A. (2025). La desinformación como amenaza híbrida: respuestas institucionales de la Unión Europea. Behavior & Law Journal, 11(1), 32-42. https://behaviorandlawjournal.com/BLJ/article/view/125

Boháček, M., y Farid, H. (2024). Lost in translation: Lip-sync deepfake detection from audio-video mismatch. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2024): WMF. https://openaccess.thecvf.com/content/CVPR2024W/WMF/papers/Bohacek_Lost_in_Translation_Lip-Sync_Deepfake_Detection_from_Audio-Video_Mismatch_CVPRW_2024_paper.pdf

Bomassi, L. (2025). The geopolitics of multipolarity: How to counter Europe's waning relevance in Southeast Asia. European Union Institute for Security Studies. https://www.iss.europa.eu/publications/briefs/geopolitics-multipolarity-how-counter-europes-waning-relevance-southeast-asia

Check Point. (2025). Check Point response to the BreachForum post on 30 March 2025 (SK183307). https://support.checkpoint.com/results/sk/sk183307

Coalition for Content Provenance and Authenticity. (2023). C2PA Technical Specification 2.2 [Technical standard]. http://spec.c2pa.org/specifications/specifications/2.2/index.html

Colás, X. y Rojas, A. (2025). Un profesor español, entre los fugitivos más buscados de Europol por "delitos de sabotaje con finalidad terrorista" a favor de Rusia. El Mundo. https://www.elmundo.es/internacional/2025/09/12/68c45d56e9cf4a53598b45ce.html

Comisión Europea. (2022a). Plan de acción contra la desinformación. Bruselas: Servicio Europeo de Acción Exterior (SEAE). https://digital-strategy.ec.europa.eu/es/policies/online-disinformation

Comisión Europea. (2022b). Code of Practice on Disinformation 2.0. Bruselas: EuropeanCommission.https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

Cordey, S. (2019). Cyber Influence Operations: An Overview and Comparative Analysis (CSS Cyberdefense Reports). Centre for Security Studies (CSS), ETH Zürich.https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2019-10-CyberInfluence.pdf

CyberProof. (2025). Beyond the blast radius: Iran's digital retaliation expands westward. CyberProof Blog. https://www.cyberproof.com/blog/beyond-the-blast-radius-irans-digital-retaliation-expands-westward/

Cyble. (2022). Threat Actor Profile: NoName057(16). Cyble Research. https://cyble.com/threat-actor-profiles/noname05716/

DarkOwl. (2024). The Digital Economy of Disinformation on the Darknet: Controlling the narrative.https://www.darkowl.com/blog-content/the-digital-economy-of-disinformation-on-the-darknet-controlling-the-narrative/

Degli-Esposti, S., y Arroyo, D. (2021). Trustworthy humans and machines. In Trust and transparency in an age of surveillance (p. 201). Routledge. https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781003120827-15/trustworthy-humans-machines-sara-degli-esposti-david-arroyo

Departamento de Seguridad Nacional. (2021). Estrategia Nacional de Seguridad 2021. Presidencia del Gobierno de España. https://www.dsn.gob.es/index.php/es/publicaciones/estrategia-de-seguridad-nacional-2021

Departamento de Seguridad Nacional. (2023). Capítulo 3. En Trabajos del Foro contra las campañas de desinformación en el ámbito de la Seguridad Nacional: Trabajos 2023. Presidencia del Gobierno de España. https://www.dsn.gob.es/es/publicaciones/otras-publicaciones/Foro-desinformacion-ambitoSN-trabajos2023

Departamento de Seguridad Nacional. (2024). Capítulo 4. En Trabajos del Foro contra las Campañas de Desinformación – Iniciativas 2024. Presidencia del Gobierno de España. https://www.dsn.gob.es/es/publicaciones/otras-publicaciones/trabajos-foro-contra-campanas-desinformacion-iniciativas-2024

ENISA – European Union Agency for Cybersecurity. (2024). ENISA Threat Landscape 2024. Atenas: Unión Europea. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024

England, A. (2025). Western nations warn of growing threat from Iranian intelligence. Financial Times. https://www.ft.com/content/60efaa9f-4c6b-4e70-98a6-70a85e-87d25a

ESRT. (2025). EE.UU. vuelve a amenazar a España si no acata "los compromisos de la OTAN". https://esrt.space/actualidad/556534-eeuu-nuevo-espana-graves-compromisos-otan

Esteban, M., y Armanini, U. (2023). La política informal de España hacia China: un enfoque coherente y europeísta. Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/la-politica-informal-de-espana-hacia-china-un-enfoque-coherente-y-europeista/

EU vs Disinfo. (2023). Narrativas rusas de desinformación sobre las sanciones. https://euvsdisinfo.eu/es/las-sanciones-no-funcionan-narrativas-rusas-de-desinformacion-sobre-las-sanciones-en-la-ue-en-ucrania-y-en-la-propia-rusia/

European External Action Service. (2025). 3rd EEAS Report on Foreign Information Manipulation and Interference (FIMI) Threats.https://www.eeas.europa.eu/eeas/3rd-eeas-report-foreign-information-manipulation-and-interference-threats-0_en

European External Action Service. (2024a, enero). 2nd report on FIMI threats (p.13).https://www.eeas.europa.eu/sites/default/files/documents/2024/EEAS-2nd-Report%20on%20FIMI%20Threats-January-2024_0.pdf

European External Action Service. (2024b, junio). Technical report on FIMI threats: Doppelganger strikes back: FIMI activities in the context of the EE24. https://euvsdisinfo.eu/uploads/2024/06/EEAS-TechnicalReport-DoppelgangerEE24_June2024.pdf

European External Action Service. (2023). 1st EEAS report on foreign information manipulation and interference threats (pp. 9–10). https://www.eeas.europa.eu/sites/default/files/documents/2023/EEAS-DataTeam-ThreatReport-2023.pdf

European External Action Service. (2022). 2022 Report on EEAS activities to counter foreign information manipulation and interference (FIMI) (pp. 5–6). https://euhybnet.eu/wp-content/uploads/2022/11/EEAS-AnnualReport-WEB_v3.4.pdf

Europol. (2020). Catching the virus: Cybercrime, disinformation and the COVID-19 pandemic.https://www.europol.europa.eu/cms/sites/default/files/documents/catching_the_virus_cybercrime_disinformation_and_the_covid-19_pandemic_0.pdf

FinCrime Central. (2024). How gift cards, loyalty and mileage programs fuel money laundering.https://fincrimecentral.com/gift-cards-mileage-programs-money-laundering/

Franceschi-Bicchierai, L. (2016). How hackers plant false flags to hide their real identities. VICE. https://www.vice.com/en/article/how-hackers-plant-false-flags-to-hide-their-real-identities/

Freedom House. (2022). Spain: Beijing's Global Media Influence 2022. https://freedomhouse.org/country/spain/beijings-global-media-influence/2022

Gatewatcher. (2025). Data breach: the operations of "Charming Kitten" revealed. https://www.gatewatcher.com/en/lab/data-breach-the-operations-of-charming-kitten-revealed/

González, M. (2025). El informe de Seguridad Nacional atribuye a Rusia campañas desestabilizadoras de desinformación por la DANA. El País. https://elpais.com/espana/2025-05-22/el-informe-de-seguridad-nacional-atribuye-a-rusia-campanas-de-desinformacion-por-la-dana.html

Global Influence Operations. (2025). Israel-Iran conflict and AI disinformation: 100 million views on fake content. https://www.global-influence-ops.com/israel-iran-conflict-ai-disinformation-100m-views-on-fake-content/

Govorit Moskva. (2025). В Испании прошли многотысячные протесты против роста цен на жильё [45 000 españoles se manifestaron contra la amnistía a los partidarios de la independencia de Cataluña]. https://govoritmoskva.ru/news/396540/

Group-IB. (2025). APAC Intelligence Insights Report. https://go.group-ib.com/hubfs/report/protected/group-ib-intelligence-report-may-2025-apac.pdf

Guancha. (2025a, 6 de mayo). 4月西班牙电车销量猛增，特斯拉大幅下滑. [Abril registra un fuerte aumento en las ventas de tranvías en España, mientras que las ventas de Tesla caen significativamente]. https://www.guancha.cn/qiche/2025_05_06_774762.shtml

Guancha. (2025b, 13 de agosto). 西班牙是个有趣的案例，在防务和中国问题上对抗特朗普还能免于报复. [España presenta un caso de estudio intrigante: puede desafiar a Trump en materia de defensa y China sin incurrir en represalias]. https://www.guancha.cn/internation/2025_08_13_786393.shtml

Guancha. (2025c, 30 de agosto). 美欧持续重压，最后一刻西班牙取消华为合同. [En medio de la presión sostenida de Estados Unidos y Europa, España canceló su contrato con Huawei en el último momento]. https://www.guancha.cn/internation/2025_08_30_788267.shtml

Hakala, J., y Melnychuk, J. (2021). Russia's strategy in cyberspace. NATO Strategic Communications Centre of Excellence. https://stratcomcoe.org/publications/russias-strategy-in-cyberspace/210

Hamayun, M. (2025). Generative AI and Agentic Systems: The New Frontline in Phishing and Smishing Defense. Check Point Blog. https://blog.checkpoint.com/executive-insights/generative-ai-and-agentic-systems-the-new-frontline-in-phishing-and-smishing-defense/

Hoeren, T. (2025). China and Spain bridge data worlds, build trust. China Daily. https://global.chinadaily.com.cn/a/202507/26/WS68843b98a310ad07b5d9222d.html

Hu, L., Mohamed, N., Melicher, B., Starov, A., Tong, H., Wilhoit, K., y Farooqi, S. (2024). The emerging dynamics of deepfake scam campaigns on the web. Unit 42, Palo Alto. https://unit42.paloaltonetworks.com/dynamics-of-deepfake-scams/

Ifeng News. (2025). 美欧持续重压，最后一刻西班牙“毁约”. [En medio de la presión sostenida de Estados Unidos y Europa, España renegó del acuerdo en el último momento]. https://news.ifeng.com/c/8mEkEDfx0fM

Insikt Group. (2023). Empire Dragon accelerates covert information operations, converges with Russian narratives. Recorded Future. https://www.recordedfuture.com/research/empire-dragon-accelerates-covert-information-operations-converges-russian-narratives

Insikt Group. (2024a). “Operation Undercut” shows multifaceted nature of SDA’s influence operations. Recorded Future. https://www.recordedfuture.com/research/operation-undercut-shows-multifaceted-nature-sdas-influence-operations

Insikt Group. (2024b). Targets, objectives, and emerging tactics of political deepfakes. Recorded Future. https://www.recordedfuture.com/research/targets-objectives-emerging-tactics-political-deepfakes

Intel471. (2025). Pro-Russian hacktivism: Shifting alliances, new groups and risks. https://www.intel471.com/blog/pro-russian-hacktivism-shifting-alliances-new-groups-and-risks

International Republican Institute. (2023). Countering China’s information manipulation: A toolkit for understanding and action. https://www.iri.org/wp-content/uploads/2023/09/Web_IRI-Toolkit-Building-Resilience-to-PRC-Information-Manipulation.pdf

Kalajdziovski, N., & Collard, S. (2025). Disinformation: The cyber threat hiding in plain sight. Security Alliance Blog. https://www.secalliance.com/blog/disinformation-the-cyber-threat-hiding-in-plain-sight

Kia, S. (2025). Iran News: Spanish outlet reports European probe into regime’s attack on Dr. Vidal-Quadras. NCRI, National Council of Resistance of Iran. https://www.ncr-iran.org/en/news/terrorism-a-fundamentalism/iran-news-spanish-outlet-reports-european-probe-into-regimes-attack-on-dr-vidal-quadras/

Lakshmanan, R. (2025). Iranian APT35 hackers targeting Israeli institutions with sophisticated phishing attacks. The Hacker News. https://thehackernews.com/2025/06/iranian-apt35-hackers-targeting-israeli.html

Lee, M. (2025). US and NATO allies warn of increasing Iranian threats in Europe, North America. AP News. https://apnews.com/article/us-nato-iran-3677783597cd-82ba16df02537083bf65

Li, Y., Yang, X., Sun, P., Qi, H., & Lyu, S. (2020). Celeb-DF: A large-scale challenging dataset for DeepFake forensics (arXiv preprint arXiv:1909.12962). https://arxiv.org/abs/1909.12962

Lupton, D. (2019). Data selves: More-than-human perspectives. Cambridge: Polity Press. https://www.politybooks.com/bookdetail?book_slug=data-selves-more-than-human-perspectives—9781509536412

Maldita. (2025). Nearly 300 contents published in five days: how the Russian network Pravda takes advantage of the Torre Pacheco riots and spreads disinformation. https://maldita.es/investigaciones/20250717/300-publications-pravda-torre-pacheco/

Marjanov, T., & Hutchings, A. (2025). SoK: Digging into the Digital Underworld of Stolen Data Markets. In 2025 IEEE Symposium on Security and Privacy (SP), (pp. 1-18). IEEE. https://ieeexplore.ieee.org/abstract/document/11023368

Márquez de la Rubia, F. (2025). La batalla por la supremacía tecnológica: EE. UU. vs. China (p. 7). Ministerio de Defensa. https://www.defensa.gob.es/documents/2073105/2392118/la_batalla_por_la_supremacia_tecnologica_2025_dieeea23.pdf

Masada, S. (2025). Desarticulación de Lumma Stealer: Microsoft lidera una acción global contra la herramienta de ciberdelincuencia. Microsoft. https://news.microsoft.com/es-es/2025/05/21/desarticulacion-de-lumma-stealer-microsoft-lidera-una-accion-global-contra-la-herramienta-de-ciberdelincuencia/

McIntosh, T., Susnjak, T., Liu, T., Xu, D., Watters, P., Liu, D., Hao, Y., Ng, A., & Halgamuge, M. (2024). Ransomware reloaded: Re-examining its trend, research and mitigation in the era of data exfiltration. ACM Computing Surveys, 57(1), 1–40. https://dl.acm.org/doi/10.1145/3691340

Meidan, O. (2025a). Keymous+: A new hacktivist collective or a DDoS-as-a-service brand in disguise?. Radware. https://www.radware.com/blog/threat-intelligence/keymous-plus-a-new-hacktivist-collective-or-a-ddos-as-a-service-brand/

Meidan, O. (2025b). Mr Hamza's Abyssal DDoS. Radware Threat Intelligence. Radware. https://www.radware.com/blog/threat-intelligence/mr-hamza-s-abyssal-ddos/

Microsoft. (2023a). Microsoft threat actor naming taxonomy. Microsoft Learn. https://learn.microsoft.com/en-us/unified-secops/microsoft-threat-actor-naming

Microsoft. (2023b, 14 de septiembre). Peach Sandstorm password spray campaigns enable intelligence collection at high-value targets. Microsoft Security Blog. https://www.microsoft.com/en-us/security/blog/2023/09/14/peach-sandstorm-password-spray-campaigns-enable-intelligence-collection-at-high-value-targets/

Microsoft. (2024a, 30 de octubre). Russia, Iran, and China continue influence campaigns in final weeks before Election Day 2024. https://www.microsoft.com/en-gb/security/security-insider/intelligence-reports/russia-iran-and-china-continue-influence-campaigns-in-final-weeks-before-election-day-2024

Microsoft. (2024b, 28 de agosto). Peach Sandstorm deploys new custom Tickler malware in long-running intelligence-gathering operations. Microsoft Security Blog. www.microsoft.com/en-us/security/blog/2024/08/28/peach-sandstorm-deploys-new-custom-tickler-malware-in-long-running-intelligence-gathering-operations/

Microsoft. (2024c, 14 de febrero). Staying ahead of threat actors in the age of AI. Microsoft Security Blog. https://www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai/

Mishra, S. (2025). Misinformation and disinformation in the digital age: A rising risk for business and investors. https://corpgov.law.harvard.edu/2025/05/12/misinformation-and-disinformation-in-the-digital-age-a-rising-risk-for-business-and-investors/

MISP Project. (s. f.). MISP – Open Source Threat Intelligence Platform & Open Standards for Threat Information Sharing. https://www.misp-project.org/

Moreno-Castro, C.; Baldi, V.; Azurmendi, A.; Paisana, M.; Iranzo-Cabrera, M.; Calvo, D.; Crespo, M.; Cabrera, Y.; Llorca-Abad, G.; Cardoso, G.; Hernández, P.; and Salaverría, R. (2023). IBERIFIER. Report on political and legal aspects of disinformation in Portugal and Spain. Pamplona: IBERIFIER. https://iberifier.eu/2023/10/20/iberifier-reports-legal-and-political-aspects-of-disinformation-in-portugal-and-spain-october-2023/

Moreno, J. (2024). La campaña propagandística de China en español no impresiona, según un informe. Voz de América. https://www.vozdeamerica.com/a/la-campana-propagandistica-de-china-en-espanol-no-impresiona-segun-un-informe/7490368.html

Newman, L.H. (2024). Notorious Iranian hackers have been targeting the space industry with a new backdoor. Wired. https://www.wired.com/story/iran-peach-sandworm-tickler-backdoor/

News-Kiev. (2025). Взрыв и пожар на заводе боеприпасов Rheinmetall в Испании: опять совпадение? [Explosión e incendio en la fábrica de municiones de Rheinmetall en España: ¿otra coincidencia?]. https://news-kiev.ru/society/2025/01/31/552871.html

Pashin, V. (2025, 5 de abril). В Испании десятки тысяч вышли на протесты против непомерного роста цен на жильё [En España decenas de miles salieron a protestar contra el desmesurado aumento de los precios de la vivienda]. RU News24. https://runews24.ru/world/05/04/2025/v-ispanii-desyatki-tyisyach-vyishli-na-protestyi-protiv-nepomernogo-rosta-czen-na-zhile

People's Daily. (2025a, 14 de enero). El comercio entre China y la UE aumenta un 1,6% en 2024, en gran medida resistente a pesar de algunas disputas comerciales: Administración General de Aduanas. http://spanish.people.com.cn/n3/2025/0114/c31620-20265965.html

People's Daily. (2025b, 7 de marzo). Empresas españolas apuestan por China y su vasto potencial de mercado. http://spanish.peopledaily.com.cn/n3/2025/0307/c31620-20286576.html

People's Daily. (2025c, 21 de julio). Embajada de China en España ataca la interferencia de Estados Unidos en las operaciones normales de Huawei en España. https://spanish.people.com.cn/n3/2025/0721/c31618-20342960.html

Privacy International. (2021). The UN Report on Disinformation: a role for privacy. https://privacyinternational.org/news-analysis/4515/un-report-disinformation-role-privacy

Rachel James. (2024). Threat-Actors-use-of-Artificial-Intelligence [Repositorio GitHub]. GitHub. https://github.com/cybershujin/Threat-Actors-Use-of-Artificial-Intelligence

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). (2022). Official Journal of the European Union, L 277, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). (2024). *Official Journal of the European Union, L *, 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Reporters Without Borders (RSF). (2025). Threat looms over Narges Mohammadi amid ongoing intimidation by state-affiliated actors. RSF. https://rsf.org/en/threat-looms-over-narges-mohammadi-amid-ongoing-intimidation-state-affiliated-actors

Reuters. (2025). Spanish court says attempt on former politician's life linked to opposition to Iran leadership. https://www.reuters.com/world/spanish-court-says-attempt-former-politicians-life-linked-opposition-iran-2025-07-09/

Ribeiro, A. (2025). Critical infrastructure warned of rising Iranian cyber threats; urged to detect, disconnect vulnerable OT, ICS devices. Industrial Cyber. https://industrialcyber.co/critical-infrastructure/critical-infrastructure-warned-of-rising-iranian-cyber-threats-urged-to-detect-disconnect-vulnerable-ot-ics-devices/

RT. (2025b, 29 de abril). Spain opens probe into major power grid failure. https://www.rt.com/news/616491-spain-blackout-cyberattack-investigation/

Sha, T. (2025). Rise in digital influence operations: An analysis into Iranian disinformation campaigns. The New Global Order. https://thenewglobalorder.com/world-news/rise-in-digital-influence-operations-an-analysis-into-iranian-disinformation-campaigns/

Shea, J. (2025). Spain set to miss NATO target. China Daily. https://global.chinadaily.com.cn/a/202502/19/WS67b53127a310c240449d6009.html

SOCRadar. (2023). Dark Web Profile: NoName057(16). SOCRadar Labs. https://socradar.io/dark-web-profile-noname05716/

SOCRadar. (2025). Dark Web Profile: Mr Hamza. SOCRadar Labs. https://socradar.io/dark-web-profile-mr-hamza/

Southcn News. (2025). [Artículo sobre Huawei y geopolítica]. https://news.southcn.com/node_179d29f1ce/a747dac6ea.shtml

Spain-China Foundation. (2025). La llegada del Año Nuevo Chino se vive en Madrid y Barcelona. https://spain-china-foundation.org/patronos/la-llegada-del-ano-nuevo-chino-se-vive-en-madrid-y-barcelona/

Spanish News-Pravda. (2025). Proveedor de APU en llamas: una fábrica Española de municiones se incendió después de una explosión. Red Pravda. https://spanish.news-pravda.com/world/2025/01/31/349384.html

Szakács, J., & Bognár, É. (2021, June). The impact of disinformation campaigns about migrants and minority groups in the EU (In-depth analysis, PE 653.641). Policy Department for External Relations, Directorate-General for External Policies, European Parliament. https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/653641/EXPO_IDA(2021)653641_EN.pdf

TopWar. (2025a, 2 de abril). Русские танки не дойдут до Пиренеев: Евросоюз требует от Испании увеличить военные расходы. [Los tanques rusos no llegarán a los Pirineos: La UE exige a España aumentar el gasto militar]. https://topwar.ru/262333-russkie-tanki-ne-dojdut-do-pireneev-evrosojuz-trebuet-ot-ispanii-uvelichit-voennye-rashody.html

TopWar. (2025b, 30 de abril). Удар по «зелёной повестке» премьера Санчеса: испанские энергетики заявили, что причиной массового сбоя могла стать перегрузка из-за всплеска солнечной генерации. [Golpe a la agenda verde del presidente Sánchez: técnicos energéticos afirman que la causa del apagón pudo ser una sobrecarga por el aumento de la generación solar]. https://topwar.ru/263867-udar-po-zelenoj-povestke-premera-sanchesa-ispanskie-jenergetiki-zajavili-chto-prichinoj-masshtabnogo-sboja-mogla-stat-peregruzka-iz-za-vspleska.html

The Diplomat in Spain. (2023). Investigation into whether the shooting of Vidal-Quadras is linked to Iran. https://thediplomatinspain.com/en/2023/11/10/investigation-into-whether-the-shooting-of-vidal-quadras-is-linked-to-iran/

The Economist. (2025a, 29 de mayo). The Uber of the underworld. https://www.economist.com/international/2025/05/29/the-uber-of-the-underworld

The Economist. (2025b, 6 de febrero). The vast and sophisticated global enterprise that is Scam Inc. https://www.economist.com/leaders/2025/02/06/the-vast-and-sophisticated-global-enterprise-that-is-scam-inc

ThreatConnect. (2016). Hacktivists vs Faketivists: Fancy Bears in Disguise. https://threatconnect.com/blog/faketivist-vs-hacktivist-how-they-differ-2/

Trilateral Research. (2025). Understanding foreign information manipulation and interference. Trilateral Research. https://trilateralresearch.com/ethics-and-human-rights/understanding-foreign-information-manipulation-and-interference

Trujillo, A., Fagni, T., & Cresci, S. (2025). The DSA transparency database: Auditing self-reported moderation actions by social media. Proceedings of the ACM on Human-Computer Interaction, 9(2), 1–28. https://dl.acm.org/doi/10.1145/3711085

T-Sanct Technologies Pvt Ltd. (2025). Emerging State-Sponsored Cyber Operations & Disinformation: Report 2024-2025. FalconFeeds. https://falconfeeds.io/reports/emerging-state-sponsored-cyber-operations-2024-2025

Van der Linden, S. (2023). Psychological inoculation against misinformation. Journal of Neurology, Neurosurgery & Psychiatry, 94(12), e2.40. https://jnnp.bmj.com/content/94/12/e2.40

World Economic Forum. (2024). The Global Risks Report 2024. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

Wright, D. (Ed.). (2025). Foreign information manipulation and interference: Case studies from the ATHENA project. Springer. https://doi.org/10.1007/978-3-031-91299-3

YaOstrov. (2025). Удар по «зелёной повестке» Санчеса: испанские энергетики заявили, что причиной сбоя стала перегрузка из-за солнечной генерации. [Golpe a la agenda verde de Sánchez: energéticos españoles afirman que la causa del apagón fue una sobrecarga por generación solar]. https://yaostrov.ru/exclusive/147118-udar-po-zelenoj-povestke-premera-sanchesa-ispanskie-jenergetiki-zajavili-chto-prichinoj-masshtabnogo-sboja-mogla-stat-peregruzka-iz-za-vspleska-solnechnoj-generacii.html

Yang, E. (2025a, 14 de marzo). Sánchez reafirma el compromiso de España con el aumento del gasto en defensa sin concretar plazos. China.org.cn. http://spanish.china.org.cn/txt/2025-03/14/content_117766483.htm

Yang, E. (2025b, 10 de septiembre). Cómo los aires acondicionados chinos están enfriando el sofocante calor de España y el resto de Europa. China.org.cn. http://spanish.china.org.cn/txt/2025-09/10/content_118073235.htm

Yasur, N., y Citrinowicz, D. (2024). Iranian foreign information manipulation and interference during the Swords of Iron war. Institute for National Security Studies (INSS). https://www.inss.org.il/publication/iran-influence/

ZakonVremeni. (s. f.). В Испании прошли многотысячные протесты против роста цен на жильё [En España se celebraron protestas multitudinarias contra la subida de los precios de la vivienda]. https://zakonvremeni.ru/news/14-4-/67879-v-ispanii-proshli-mnogotysyachnye-protesty-protiv-rosta-tsen-na-zhile.html

Zheng, T. (2025). Comercio entre China y la UE aumenta en 2024. China.org.cn. http://spanish.china.org.cn/txt/2025-01/15/content_117665647.htm

ANEXOS

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0002: Facilitate State Propaganda	Difusión sistemática de narrativas favorables al Estado para legitimar políticas y moldear percepción pública.
T0003: Leverage Existing Narratives	Reutilizar marcos narrativos ya presentes en la audiencia para aumentar credibilidad y velocidad de propagación.
T0004: Develop Competing Narratives	Crear versiones alternativas que compitan con relatos críticos para confundir al público y desplazar la agenda.
T0010: Cultivate Ignorant Agents	Instrumentalizar simpatizantes o creadores con escasa conciencia del patrocinio para amplificar mensajes.
T0014: Prepare Fundraising Campaigns	Diseño y uso de mecanismos de financiación (donaciones, criptomonedas, crowdfunding) para sostener operaciones.
T0015: Create Hashtags and Search Artifacts	Generación/uso de hashtags y artefactos de búsqueda para posicionar contenido y facilitar descubrimiento.
T0016: Create Clickbait	Titulación sensacionalista orientada a maximizar clics y alcance, a menudo en redes sociales.
T0022: Leverage Conspiracy Theory Narratives	Ampliar o explotar teorías conspirativas para sembrar desconfianza y polarizar.
T0023: Distort Facts	Reformulación o descontextualización de información verídica para producir conclusiones engañosas.
T0042: Seed Kernel of Truth	Introducir un elemento verídico como ancla para legitimar luego afirmaciones falsas.
T0044: Seed Distortions	Plantar informaciones distorsionadas intencionadamente para condicionar interpretación posterior.
T0045: Use Fake Experts	Crear o promocionar “expertos” falsos para dar apariencia técnica a la narrativa.

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0057: Organize Events	Planificación y promoción de eventos físicos o simbólicos para traducir influencia online a presencia real.
T0059: Play the Long Game	Estrategia de persistencia: mantener narrativas y activos en el tiempo para normalizar mensajes.
T0060: Continue to Amplify	Seguimiento y re-amplificación sostenida de narrativas para mantener visibilidad.
T0066: Degrade Adversary	Campañas dirigidas a deteriorar la imagen y la capacidad de actuación de actores adversos.
T0068: Respond to Breaking News Event or Active Crisis	Aprovechar crisis o noticias de última hora para insertar narrativas antes de que existan verificaciones.
T0072: Segment Audiences	Desagregación del público por variables demográficas/psicográficas para adaptar mensajes.
T0073: Determine Target Audiences	Identificación deliberada de colectivos prioritarios (ej. diásporas, grupos ideológicos).
T0074: Determine Strategic Ends	Definición de fines estratégicos (ventaja geopolítica, económica, ideológica) que guían la operación.
T0075: Dismiss	Responder desacreditando críticas: describir a los críticos como parciales o mal informados.
T0076: Distort	Cambiar marcos y contexto para alterar la interpretación de sucesos.
T0077: Distract	Desviar la atención hacia asuntos alternativos para minimizar foco sobre asuntos sensibles.
T0078: Dismay	Intimidación o amenazas dirigidas a periodistas/ actores para inhibir la crítica.

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0079: Divide	Estrategias destinadas a fracturar cohesión social entre subgrupos (polarización deliberada).
T0080: Map Target Audience Information Environment	Análisis del ecosistema informativo (tráfico, hashtags, plataformas) para optimizar inserciones.
T0081: Identify Social and Technical Vulnerabilities	Detección de puntos débiles sociales (prejuicios, cámaras de eco) y técnicos (vulnerabilidades en webs).
T0082: Develop New Narratives	Generación intencional de marcos narrativos originales para sostener operaciones.
T0083: Integrate Target Audience Vulnerabilities into Narrative	Incorporar debilidades específicas del público (miedos, intereses) para aumentar persuasión.
T0084: Reuse Existing Content	Reaprovechamiento o plagio de contenidos previos para acelerar producción y difusión.
T0085: Develop Text-Based Content	Producción de artículos, informes o posts (incl. IA) que sirven de columna vertebral narrativa.
T0086: Develop Image-based Content	Creación/edición de imágenes y memes (*deepfakes/ cheap fakes*) para soporte visual de la narrativa.
T0087: Develop Video-based Content	Producción/alteración de vídeo (incl. IA) para aumentar impacto emocional y viralidad.
T0091: Recruit Malign Actors	Captación de colaboradores, contratistas o simpatizantes para ejecutar o amplificar operaciones.
T0092: Build Network	Construcción de redes mixtas (cuentas auténticas/ inauténticas) que interactúan para generar legitimidad percibida.
T0093: Acquire/Recruit Network	Compra o financiación de redes, *botnets o proxies* para amplificación y anonimato.

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0095: Develop Owned Media Assets	Crear y gestionar medios propios (blogs, webs, canales) para controlar la narrativa.
T0096: Leverage Content Farms	Subcontratar o establecer fábricas de contenido a escala para producir y diseminar material.
T0097: Present Persona	Construcción de identidades (expertos, periodistas, ONGs) para operar en diferentes plataformas.
T0098: Create Inauthentic News Sites	Montaje de portales que simulan medios independientes para dar apariencia periodística.
T0100: Co-opt Trusted Sources	Captar o corromper fuentes de confianza (*influencers*, líderes locales) para legitimar contenidos.
T0101: Create Localized Content	Adaptar lenguaje, referencias y formatos culturales para comunidades locales o diásporas.
T0102: Leverage Echo Chambers/Filter Bubbles	Aprovechar o crear cámaras de eco para reforzar mensajes sin exposición crítica.
T0110: Formal Diplomatic Channels	Uso de canales oficiales (diplomacia, comunicados) para replicar o respaldar narrativas.
T0115: Post Content	Publicación directa en activos propios (post, artículo, canal) para fijar la narrativa.
T0116: Comment or Reply on Content	Insertar comentarios y réplicas (a veces automatizadas) para dar la impresión de conversación y apoyo.
T0117: Attract Traditional Media	Diseñar contenido y señuelos para que medios convencionales repliquen la narrativa.
T0118: Amplify Existing Narrative	Reforzar historias ya virales mediante reposts coordinados y cobertura cruzada.

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0119: Cross-Posting	Publicar el mismo mensaje en múltiples plataformas/ grupos para dar apariencia de consenso.
T0123: Control Information Environment through Offensive Cyberspace Operations	Uso de operaciones técnicas (*phishing*, DDoS, intrusión) para alterar disponibilidad y visibilidad de información.
T0126: Encourage Attendance at Events	Llamadas y facilidades para que audiencias asistan a actos que legitiman la narrativa.
T0127: Physical Violence	Cuando corresponde, fomento u organización de violencia física como recurso intimidatorio o disuasorio.
T0128: Conceal People	Ocultación de identidades y orígenes de cuentas/ activistas para evitar atribución y sanción.
T0129: Conceal Operational Activity	Técnicas para borrar huellas operativas, romper vínculos y falsear orígenes.
T0133: Measure Effectiveness	Monitorización de métricas (alcance, *engagement*) para evaluar impacto y ajustar tácticas.
T0134: Message Reach / Social Media Engagement	Seguimiento específico del alcance y la participación para priorizar plataformas y narrativas.
T0135: Undermine	Campañas para socavar confianza en instituciones o actores (difamación, subversión).
T0136: Cultivate Support	Estrategias para generar simpatía y respaldo (justificaciones, exaltación, reclutamiento).
T0137: Make Money	Técnicas específicas orientadas a la obtención de beneficios (*scams*, extorsión, monetización).
T0138: Motivate to Act	Mensajes orientados a inducir comportamientos (*encouragement/provoke/compel*).

Identificación y descripción de las TTP

TTP	DESCRIPCIÓN
T0139: Dissuade from Acting	Mensajes y operaciones que intentan impedir contra-acciones (intimidación, desaliento).
T0140: Cause Harm	Acciones destinadas a causar daño reputacional, económico o físico a objetivos.
T0143: Legitimise Persona	Dar apariencia de autenticidad a perfiles (creación de historial, pruebas cruzadas).
T0144: Persona Legitimacy Evidence	Crear evidencia (presencia multiplataforma, publicaciones antiguas) que respalde la persona fabricada.
T0145: Establish Account Imagery	Uso de imágenes (IA, stock, atractivas) para perfilar cuentas con mayor atracción y credibilidad.
T0146: Account Asset	Creación y gestión de cuentas (verificadas, automatizadas, *lookalikes*) como activos operativos.
T0147: Software Asset	Desarrollo o uso de herramientas y apps (incl. plataformas móviles) que facilitan la difusión.
T0148: Financial Instrument	Uso de instrumentos y plataformas financieras (*exchanges*, *wallets*, *e-commerce*) para mover fondos y monetizar.
T0150: Asset Origin (Bulk / Dormant / Rented / Reused)	Diversas prácticas sobre el origen de activos (masivos, durmientes, comprados, alquilados, reutilizados).
T0151 / T0152 / T0153 / T0154: Platform & Delivery Assets	Selección y uso de plataformas (social, hosting, *ads*, LLM/IA generativa, algoritmos de recomendación) como infraestructura de entrega.
T0155: Gated Asset / Paid Access / Encrypted Channels	Activos con acceso controlado (*membership*, pago, canales cifrados) para generar exclusividad y confianza.

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0074: Determine Strategic Ends				
T0074.002: Domestic Political Advantage	X	X	1	3
T0074.003: Economic Advantage	X		2	3
T0074.001: Geopolitical Advantage		X	1	2
T0074.004: Ideological Advantage		X	1	2
T0073: Determine Target Audiences	X	X	1	3
T0140: Cause Harm				
T0140.001: Defame	X	X	2	4
T0140.002: Intimidate	X	X	2	4
T0136: Cultivate Support				
T0136.004: Boost Reputation	X		2	3
T0136.006: Cultivate Support for Ally		X	1	2
T0136.005: Cultvate Support for Initiative			2	2
T0136.001: Defend Reputaton			2	2
T0136.003: Energise Supporters	X		1	2
T0136.008: Increase Prestige	X		2	3
T0136.002: Justify Action		X	1	2
T0136.007: Recruit Members		X	1	2
T0066: Degrade Adversary		X	3	4
T0078: Dismay			1	1
T0075: Dismiss				
T0075.001: Discredit Credible Sources		X	2	1
T0139: Dissuade from Acting				
T0139.003: Deter			1	1
T0139.001: Discourage	X		1	1
T0139.002: Silence			1	1

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0076: Distort			1	1
T0077: Distract		X	1	2
T0079: Divide			2	2
T0002: Facilitate State Propaganda			2	2
T0137: Make Money				
T0137.005: Extort	X		0	1
T0137.003: Raise Funds		X	0	1
T0137.002: Scam	X		0	1
T0138: Motivate to Act				
T0138.003: Compel			1	1
T0138.001: Encourage		X	1	2
T0138.002: Provoke		X	1	2
T0135: Undermine				
T0135.004: Polarise		X	2	3
T0135.001: Smear			2	2
T0135.003: Subvert			1	1
T0135.002: Thwart			1	1
T0081: Identify Social and Technical Vulnerabilities				
T0081.001: Find Echo Chambers	X		1	2
T0081.002: Identify Data Voids			1	1
T0081.005: Identify Existing Conspiracy Narratives/ Suspicions			1	1
T0081.004: Identify Existing Fissures			1	1
T0081.003: Identify Existing Prejudices		X	2	3
T0081.008: Identify Media System Vulnerabilities			1	1
T0081.007: Identify Target Audience Adversaries		X	2	3

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0081.006: Identify Wedge Issues		X	1	2
T0080: Map Target Audience Information Environment				
T0080.005: Assess Degree/Type of Media Access	X		1	2
T0080.004: Conduct Web Traffic Analysis			1	1
T0080.002: Evaluate Media Surveys			1	**1**
T0080.003: Identify Trending Topics/Hashtags	X		1	2
T0080.001: Monitor Social Media Analytics			1	**1**
T0072: Segment Audiences				
T0072.002: Demographic Segmentation	X		1	**2**
T0072.003: Economic Segmentation			1	**1**
T0072.001: Geographic Segmentation		X	1	**2**
T0072.005: Political Segmentation		X	1	**2**
T0072.004: Psychographic Segmentation	X		1	**2**
T0082: Develop New Narratives	X		0	**1**
T0083: Integrate Target Audience Vulnerabilities into Narrative	X		2	**3**
T0022: Leverage Conspiracy Theory Narratives				
T0022.001: Amplify Existing Conspiracy Theory Narratives			1	**1**
T0022.002: Develop Original Conspiracy Theory Narratives			1	**1**
T0003: Leverage Existing Narratives	X	X	1	**3**
T0068: Respond to Breaking News Event or Active Crisis	X	X	3	**5**
T0015: Create Hashtags and Search Artefacts				
T0015.002: Create New Hashtag	X		1	**2**
T0015.001: Use Existing Hashtag	X		1	**2**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0086: Develop Image-Based Content				
T0086.003: Deceptively Edit Images (Cheap Fakes)	X	X	0	**2**
T0086.002: Develop AI-Generated Images (Deepfakes)	X	X	0	**2**
T0086.001: Develop Memes			1	**1**
T0085: Develop Text-Based Content				
T0085.007: Create Fake Research			1	**1**
T0085.001: Develop AI-Generated Text	X		0	**1**
T0085.004: Develop Document		X	2	**3**
T0085.003: Develop Inauthentic News Articles	X		2	**3**
T0085.006: Develop Opinion Article	X		1	**2**
T0087: Develop Video-Based Content				
T0087.002: Deceptively Edit Video (Cheap Fakes)	X	X	0	**2**
T0087.001: Develop AI-Generated Videos (Deepfakes)	X	X	0	**2**
T0023: Distort Facts				
T0023.001: Reframe Context		X	2	**3**
T0084: Reuse Existing Content				
T0084.002: Plagiarise Content	X		0	**1**
T0146: Account Asset				
T0146.007: Automated Account Asset	X		3	**4**
T0146.001: Free Account Asset			1	**1**
T0146.005: Lookalike Account ID			1	**1**
T0146.006: Open Access Platform			1	**1**
T0146.003: Verified Account Asset	X		1	**2**
T0093: Acquire/Recruit Network				
T0093.002: Acquire Botnets			2	**2**
T0093.001: Fund Proxies			2	**2**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0150: Asset Origin				
T0150.008: Bulk Created Asset			2	**2**
T0150.005: Compromised Asset			1	**1**
T0150.002: Dormant Asset			1	**1**
T0150.001: Newly Created Asset			1	**1**
T0150.003: Pre-Existing Asset			1	**1**
T0150.007: Rented Asset			1	**1**
T0150.004: Repurposed Asset			1	**1**
T0092: Build Network				
T0092.003: Create Community or Sub-Group		X	2	**3**
T0092.001: Create Organisations			2	**2**
T0092.002: Use Follow Trains			1	**1**
T0010: Cultivate Ignorant Agents			3	**3**
T0095: Develop Owned Media Assets		X	2	**3**
T0145: Establish Account Imagery				
T0145.002: AI-Generated Account Imagery	X		1	**2**
T0145.006: Attractive Person Account Imagery	X		1	**2**
T0145.005: Illustrated Character Account Imagery			1	**1**
T0145.007: Stock Image Account Imagery	X		0	**1**
T0148: Financial Instrument				
T0148.002: Bank Account Asset		X	0	**1**
T0148.006: Crowdfunding Platform		X	0	**1**
T0148.008: Cryptocurrency Exchange Platform	X		0	**1**
T0148.009: Cryptocurrency Wallet	X		0	**1**
T0148.007: eCommerce Platform	X		0	**1**
T0148.005: Subscription Processing Capability		X	0	**1**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0094: Infiltrate Existing Networks				
T0094.001: Identify Susceptible Targets in Networks			1	**1**
T0094.002: Utilise Butterfly Attacks			1	**1**
T0096: Leverage Content Farms				
T0096.001: Create Content Farms			2	**2**
T0096.002: Outsource Content Creation to External Organisations			2	**2**
T0149: Online Infrastructure				
T0149.003: Lookalike Domain			1	**1**
T0014: Prepare Fundraising Campaigns				
T0014.002: Raise Funds from Ignorant Agents		X	1	**2**
T0014.001: Raise Funds from Malign Actors			1	**1**
T0091: Recruit Malign Actors				
T0091.003: Enlist Troll Accounts			1	**1**
T0091.001: Recruit Contractors			1	**1**
T0091.002: Recruit Partisans		X	1	**2**
T0147: Software Asset				
T0147.003: Malware Asset			1	**1**
T0147.004: Mobile App Asset			1	**1**
T0100: Co-Opt Trusted Sources				
T0100.002: Co-Opt Grassroots Groups	X		1	**2**
T0100.003: Co-Opt Influencers	X		2	**3**
T0100.001: Co-Opt Trusted Individuals	X		2	**3**
T0098: Establish Inauthentic News Sites				
T0098.001: Create Inauthentic News Sites			2	**2**
T0143: Persona Legitimacy				
T0143.001: Authentic Persona			2	**2**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0143.002: Fabricated Persona			1	**1**
T0143.004: Parody Persona			1	**1**
T0144: Persona Legitimacy Evidence				
T0144.002: Persona Template			1	**1**
T0144.001: Present Persona across Platforms	X		1	**2**
T0097: Present Persona				
T0097.103: Activist Persona			1	**1**
T0097.108: Expert Persona	X		2	**3**
T0097.203: Fact Checking Organisation Persona			1	**1**
T0097.104: Hacktivist Persona			2	**2**
T0097.100: Individual Persona	X		1	**2**
T0097.102: Journalist Persona			2	**2**
T0097.101: Local Persona			1	**1**
T0097.202: News Outlet Persona			1	**1**
T0097.110: Party Official Persona			1	**1**
T0097.208: Social Cause Persona			1	**1**
T0097.204: Think Tank Persona			1	**1**
T0016: Create Clickbait			1	**1**
T0101: Create Localised Content			2	**2**
T0102: Leverage Echo Chambers/Filter Bubbles				
T0102.002: Create Echo Chambers/Filter Bubbles			1	**1**
T0102.003: Exploit Data Voids			1	**1**
T0102.001: Use Existing Echo Chambers/Filter Bubbles	X	X	1	**3**
T0151: Digital Community Hosting Asset				
T0151.005: Chat Community Server	X	X	0	**2**
T0151.004: Chat Platform	X		1	**2**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0151.014: Comments Section			1	**1**
T0151.010: Community Forum Platform	X		0	**1**
T0151.002: Online Community Group	X	X	0	**2**
T0151.015: Online Game Platform	X		0	**1**
T0151.001: Social Media Platform	X		2	**3**
T0154: Digital Content Creation Asset				
T0154.001: AI LLM Platform			1	**1**
T0154.002: AI Media Platform			1	**1**
T0153: Digital Content Delivery Asset				
T0153.006: Content Recommendation Algorithm			1	**1**
T0153.007: Direct Messaging		X	0	**1**
T0153.005: Online Advertising Platform			1	**1**
T0152: Digital Content Hosting Asset				
T0152.004: Website Asset			1	**1**
T0110: Formal Diplomatic Channels			2	**2**
T0155: Gated Asset				
T0155.003: Approval Gated Asset	X		0	**1**
T0155.007: Encrypted Communication Channel	X		0	**1**
T0155.002: Invite Gated Asset	X		0	**1**
T0029: Online Polls		X	0	**1**
T0111: Traditional Media				
T0111.002: Newspaper			1	**1**
T0044: Seed Distortions			2	**2**
T0042: Seed Kernel of Truth			1	**1**
T0020: Trial Content			1	**2**
T0045: Use Fake Experts			1	**1**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0117: Attract Traditional Media			2	**2**
T0116: Comment or Reply on Content				
T0116.001: Post Inauthentic Social Media Comment	X		1	**2**
T0115: Post Content				
T0115.003: One-Way Direct Posting		X	2	**3**
T0115.002: Post Violative Content to Provoke Takedown and Backlash			1	**1**
T0115.001: Share Memes		X	1	**2**
T0118: Amplify Existing Narrative		X	1	**2**
T0039: Bait Influencer			2	**2**
T0119: Cross-Posting				
T0119.003: Post across Disciplines			1	**1**
T0119.001: Post across Groups	X	X	1	**3**
T0119.002: Post across Platform	X	X	1	**3**
T0122: Direct Users to Alternative Platforms			1	**1**
T0049: Flood Information Space				
T0049.003: Bots Amplify via Automated Forwarding and Reposting			1	**1**
T0049.006: Conduct Keyword Squatting			1	**1**
T0049.005: Conduct Swarming			1	**1**
T0049.008: Generate Information Pollution			1	**1**
T0049.007: Inauthentic Sites Amplify News and Narratives	X		1	**2**
T0049.001: Trolls Amplify and Manipulate			1	**1**
T0049.004: Utilise Spamoflauge			1	**1**
T0120: Incentivize Sharing				
T0120.001: Use Affiliate Marketing Programmes	X		0	**1**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0123: Control Information Environment through Offensive Cyberspace Operations				
T0123.002: Block Content		X	0	**1**
T0123.001: Delete Opposing Content		X	0	**1**
T0123.003: Destroy Information Generation Capabilities		X	0	**1**
T0048: Harass				
T0048.001: Boycott/"Cancel" Opponents			1	**1**
T0048.004: Dox			1	**1**
T0048.002: Harass People Based on Identities			1	**1**
T0048.003: Threaten to Dox			1	**1**
T0126: Encourage Attendance at Events				
T0126.001: Call to Action to Attend		X	2	**3**
T0126.002: Facilitate Logistics or Support for Attendance			X	**1**
T0057: Organise Events				
T0057.002: Conduct Symbolic Action		X	2	**3**
T0057.001: Pay for Physical Action			1	**1**
T0127: Physical Violence				
T0127.001: Conduct Physical Violence			1	**1**
T0127.002: Encourage Physical Violence			1	**1**
T0128: Conceal Information Assets				
T0128.002: Conceal Network Identity			1	**1**
T0128.004: Launder Information Assets			1	**1**
T0128.001: Use Pseudonyms	X		0	**1**
T0130: Conceal Infrastructure				
T0130.005: Obfuscate Payment	X		0	**1**
T0130.004: Use Cryptocurrency	X		0	**1**

Matriz comparativa de TTP por clasificación de actores

TTP DISARM	Cibercrimen	Hacktivismo	APTs	TOTAL(APT, hacktivismo, cibercrimen)
T0129: Conceal Operational Activity				
T0129.003: Break Association with Content			1	**2**
T0129.001: Conceal Network Identity			1	**1**
T0129.005: Coordinate on Encrypted/Closed Networks	X		0	**1**
T0129.004: Delete URLs	X		0	**1**
T0129.006: Deny Involvement			2	**2**
T0129.002: Generate Content Unrelated to Narrative	X		1	**2**
T0129.010: Misattribute Activity			1	**1**
T0129.009: Remove Post Origins	X		1	**2**
T0060: Continue to Amplify			2	**2**
T0131: Exploit TOS/Content Moderation				
T0131.001: Legacy Web Content			0	**1**
T0131.002: Post Borderline Content			0	**1**
T0059: Play The Long Game			2	**2**
T0133: Measure Effectiveness				
T0133.001: Behaviour Changes			1	**1**
T0133.002: Content			1	**1**
T0134: Measure Effectiveness Indicators (or KPIs)				
T0134.001: Message Reach			1	**1**
T0134.002: Social Media Engagement			1	**1**

Detalle técnico sobre la TTP T0068 – Respond to breaking news evento or active crisis

La técnica *T0068 – Respond to Breaking News Event or Active Crisis* se fundamenta en la explotación informativa de situaciones críticas o eventos de alta exposición mediática con el fin de introducir desinformación en tiempo real. La identificación temprana de estas actividades es esencial para limitar su impacto y preservar la integridad del espacio informativo.

La detección y mitigación requieren una aproximación multimodal que combine el análisis técnico con la verificación y la comunicación pública. A continuación, se detallan las principales señales de detección y las fases metodológicas de respuesta ante incidentes de esta naturaleza.

Señales de detección:

- **Publicación reactiva inmediata:** Una de las señales más características de esta técnica es la aparición de contenido sospechoso en los primeros minutos tras un evento crítico, antes de que existan fuentes oficiales verificadas. La inmediatez de publicación suele indicar una preparación previa del material manipulado, evidenciando coordinación operativa entre actores o redes automatizadas.

- **Narrativas de alto impacto emocional:** El uso de un lenguaje polarizante, la apelación al miedo o la indignación, y las afirmaciones graves sin respaldo documental son indicadores frecuentes. Estas narrativas buscan estimular reacciones instantáneas, impulsivas y virales, aprovechando la atención mediática inicial.

- **Manipulación contextual:** Reutilización de vídeos o imágenes antiguas, modificadas o fuera de contexto que simulan declaraciones vinculados al evento.

- **Análisis de patrones de difusión:** La detección de picos anómalos de actividad, replicación simultánea de mensajes y coordinación entre cuentas de recién creación o automatizadas permite identificar la propagación coordinada en múltiples plataformas.

- **Falsas atribuciones institucionales:** Simulación de comunicados, logotipos o formatos oficiales sin trazabilidad ni metadatos válidos.

Mitigaciones clave:

- **Monitorización reactiva en tiempo real:** Despliegue de sistemas de vigilancia continua capaces de detectar contenido emergente relacionado con el evento durante sus primeras fases de difusión. Se utilizan plataformas de vigilancia digital (Brandwatch, Talkwalker, CrowdTangle) con filtros por idioma y localización, junto con algoritmos de análisis temporal que detectan picos anómalos de actividad. Los modelos multimodales como *AV-HuBERT, CLIP o Vision Temporal Transformers* permiten analizar simultáneamente texto, imagen y audio, detectando inconsistencias narrativas o simulaciones institucionales. Estos sistemas se integran en paneles operativos dentro de SOCs/CSIRTs, activando protocolos de verificación y respuesta.

- **Análisis semántico y narrativo:** Aplicación de modelos de procesamiento del lenguaje natural (*BERT, RoBERTa, GPT-based classifiers*) para identificar desviaciones respecto a fuentes oficiales, contradicciones internas y patrones lingüísticos característicos de la desinformación. Este análisis detecta títulos, subtítulos, descripciones y transcripciones, comparándolos con comunicados institucionales mediante técnicas de *semantic similarity* y *stance detection*, identificando simulaciones o distorsiones narrativas.

- **Cruce de señales contextuales:** Verificación del contenido frente a fuentes confiables y bases de datos de referencia. Se emplean sistemas de fact-checking automatizado y análisis de metadatos (fecha, formato, compresión) para detectar reutilización o manipulación. También se contrastan elementos visuales como logotipos, estilos gráficos o formatos con los utilizados oficialmente, para identificar falsificaciones.

- **Clasificación de riesgo reputacional e institucional:** Categorización del contenido según su potencial de daño, considerando el perfil de la persona afectada (CEO, ministro, portavoz), el contexto del evento (electoral, financiero, de seguridad) y el alcance proyectado. Se utilizan sistemas de scoring reputacional que combinan análisis de impacto mediático, sensibilidad institucional y probabilidad de viralización. Esta clasificación determina el nivel de escalado, la urgencia de respuesta y el tipo de mitigación a aplicar.

- **Activación de canales de escalado con plataformas:** Comunicación directa con plataformas digitales de gran escala (VLOPs y VLOSEs), solicitando medidas urgentes como etiquetado, limitación de visibilidad algorítmica (circuit breaker) o retirada inmediata del contenido mediante takedown acelerado conforme a los SLAs establecidos. Este proceso debe documentarse y mantenerse trazable para auditoría y cumplimiento. Todo el proceso debe estar alineado con los

marcos regulatorios del *DSA* y el *AI Act*, garantizando trazabilidad y documentación para auditoría. Ver mitigación reflejada en la técnica T0087.001 - Develop AI generated videos (*deepfakes*) en el **Anexo IV**.

- **Publicación de nota técnica y comunicación oficial:** Emisión de una nota técnica que documente las evidencias de manipulación (análisis forense, metadatos, comparativas con fuentes oficiales) junto con un comunicado institucional breve, verificable y coherente con la narrativa oficial. Esta nota debe acompañarse de una comunicación institucional clara que refuerce la narrativa oficial, reduzca el impacto reputacional y prevenga la consolidación del contenido falso como versión dominante.

- **Vigilancia narrativa post-evento:** Mantener la monitorización activa para detectar variantes o mutaciones del contenido, como recortes, subtitulados falsos o republicaciones. Se utilizan técnicas de *hashing perceptual*, *análisis semántico* y *seguimiento de patrones de difusión*. Esta vigilancia permite aplicar medidas de contención continuada, evitando que el contenido manipulado resurja en otros canales.

Posible caso de uso:

Tras un atentado en una capital europea, se difundió en redes un vídeo atribuido falsamente a un ministro, en el que se le veía pronunciando declaraciones supuestamente polémicas. El contenido fue publicado 12 minutos después del suceso, presentando inconsistencias entre audio y vídeo (ASR vs. lip reading) y reutilizando imágenes de archivo con compresión alterada. El sistema de vigilancia detectó el patrón anómalo de publicación y activó el protocolo de crisis:

- Se cruzó el material con fuentes oficiales, confirmando la manipulación.

- Se notificó a la plataforma correspondiente a través del canal de escalado del DSA, solicitando etiquetado urgente y limitación de visibilidad.

- Finalmente, se publicó una nota técnica con evidencia forense, acompañada de un comunicado institucional que restableció la narrativa verificada.

El contenido fue desindexado y retirado en menos de tres horas, y su evolución narrativa se monitorizó posteriormente para evitar su reaparición bajo nuevas variantes.

Detalle técnico sobre la TTP T0087.001 – *Develop AI generated* videos (*deepfakes*)

La técnica *T0087.001 - Develop AI Generated Videos (Deepfakes)* hace referencia a la manipulación o generación de vídeos mediante inteligencia artificial con el objetivo de atribuir falsamente palabras, gestos o acciones a personas reales. A diferencia de las imágenes estáticas, el vídeo combina continuidad temporal, audio y movimiento lo que otorga mayor verosimilitud al engaño y complica su detección. A continuación, se detallan las principales señales de detección y las fases metodológicas de respuesta ante incidentes de esta naturaleza.

Señales de detección

- **Desincronización entre audio y vídeo (lip-sync reenactment):** Se identifican incoherencias entre el movimiento labial y los fonemas esperados en el audio. Este tipo de simulación sincronizada de labios se detecta comparando la transcripción obtenida por *lip reading* (análisis visual de los labios en el vídeo) con la generada por sistemas *speech-to-text*. Divergencias sostenidas entre ambas transcripciones son indicativas de manipulación.

- **Patrones temporales anómalos:** Presencia de micro movimientos irregulares o "*micro jitter*" en labios, mandíbula o expresión facial, así como transiciones incoherentes entre fotogramas. Estas irregularidades se detectan mediante modelos que analizan la consistencia espacio-temporal de la secuencia audiovisual.

- **Head pose / landmarks incosistentes:** Divergencias entre la orientación de la cabeza (ángulos de yaw, pitch y roll) y la expresión facial detectada mediante puntos de referencia (landmarks); estas incoherencias pueden indicar manipulación. Se considera una señal complementaria, aunque su robustez depende del método y del dataset (colecciones de vídeos reales y falsos) utilizado para entrenar los modelos.

- **Artefactos espaciales / espectrales:** Detección de rastros propios del proceso de síntesis, como bordes irregulares, mezcla imperfecta entre regiones (*blending*), texturas anómalas o huellas de compresión no naturales. Los detectores basados en arquitecturas como las redes que analizan pixeles y filtros (*Convolutional Neural Networks* - CNN) y los modelos que capturan relaciones complejas en la imagen (Transformers) son capaces de identificar estos patrones de alta frecuencia combinando análisis visual y temporal para detectar inconsistencias en el video analizado.

- **Meta-señales de procedencia:** Ausencia, alteración o invalidez de las credenciales digitales (*Content Credentials*) o manifiestos C2PA, especialmente en material audiovisual que debería estar autenticado por fuentes institucionales o medios de comunicación.

Mitigaciones clave:

- **Modelos temporales multimodales (A/V)**: La detección de deepfakes requiere sistemas que analicen conjuntamente la consistencia temporal y la coherencia del audio-vídeo. A continuación, se enumera la aplicación técnica:

    - ***Desacoplamiento y comparación A/V.*** Se procesa por separado el audio mediante Automatic Speech Recognition (ASR) que convierte el audio en texto, mientras que *lip reading* analiza el movimiento de labios en el video; se mide la coherencia semántica y temporal entre ambos. Se activa una alerta cuando se produce una falta de sincronía superiores a ciertos umbrales (por ejemplo, mayores a X ms) o las transcripciones divergen significativamente.

    - ***Detectores espaciotemporales.*** Arquitecturas como *Vision Temporal Transformers*, fusiones multimodales o extractores *AV-HuBERT* detectan inconsistencias entre gestos y voz, especialmente en regiones faciales críticas. *Vision Temporal Transformers* son arquitecturas basadas en transformers que capturan relaciones entre fotogramas para detectar inconsistencias; las fusiones multimodales combinan señales de audio y vídeo para mejorar la precisión; *AV HuBERT* es un extractor que genera representaciones robustas de audio y vídeo para tareas de detección. Se centran en la región oral y su relación con la posición de la cabeza y gestos para identificar manipulaciones sutiles.

    - ***Priorizar detección en flujo (streaming):*** permite realizar una respuesta temprana en incidentes virales, aplicando los modelos de detección directamente sobre contenido en tiempo real, en lugar de esperar a procesar vídeos completos. Esto permite activar alertas y mitigaciones de forma inmediata cuando un deepfake comienza a viralizarse, reduciendo su alcance antes de que se consolide.

Posible caso de uso: Ante un vídeo viral sospechoso, el sistema compara ASR (audio) con *lip reading* (vídeo), detectando una falta de sincronía fonética durante 30 segundos, confirmadas por un *Vision Temporal Transformer* que marca anomalías en la región oral. Como resultado se genera una alerta, degradando el alcance del contenido en la plataforma y se efectúa un análisis forense para confirmar la manipulación.

- **Coordinación con plataformas y "*takedown*" (marco DSA/AI Act):** Reducir la exposición y frenar la amplificación algorítmica de *deepfakes* a través de procedimientos de emergencia y etiquetado o retirada en colaboración con plataformas. A continuación, se enumera la aplicación técnica asociada al caso:

    - Se activan canales de mitigación de emergencia conforme al Digital Services Act, que permiten reducir el impacto de contenidos manipulados en contextos sensibles. Esto incluye equipos dedicados para respuesta rápida, etiquetado de deepfakes para advertir a los usuarios, un mecanismo que detiene la amplificación automática en los algoritmos de recomendación para frenar la viralización (circuit breaker), y un procedimiento para retirar el contenido de forma urgente mediante protocolos priorizados (takedown). Todo debe garantizar trazabilidad con registros detallados para auditoría y cumplimiento normativo.

    - Exigir marcado y etiquetado de contenidos generados por IA cuando proceda, siguiendo las directrices del AI Act, que establece la obligación de indicar claramente si un contenido ha sido creado o manipulado mediante inteligencia artificial. Este marcado debe integrarse en las políticas de las plataformas para garantizar transparencia y trazabilidad, evitando que los usuarios confundan material sintético con contenido auténtico.

    - Mantener listas de escalado y SLAs con VLOPs/VLOSEs, estableciendo prioridades de riesgo para efectuar la comunicación y elaborar acuerdos de respuesta con plataformas digitales de gran escala (VLOPs/VLOSEs) y priorizando incidentes que afecten a figuras públicas o procesos electorales. Incluye tiempos máximos de actuación, trazabilidad y clasificación por sensibilidad del contenido.

Posible caso de uso: Se detecta un deepfake sobre un directivo antes de los resultados financieros. El sistema identifica incoherencias entre audio (ASR) y vídeo (*lip reading*), junto con anomalías espaciotemporales en la región oral mediante Vision Temporal Transformer. Se activa el protocolo DSA y se contacta con la plataforma (VLOP) a través de un canal prioritario definido en la lista de escalado, se solicita etiquetado y la retirada urgente del contenido (*takedown*) conforme al SLA acordado, se bloquea la amplificación algorítmica (*circuit breaker*), y se publica la nota técnica con las evidencias resumidas para la trazabilidad y la comunicación pública.

- **Repositorios de muestras y reentrenamiento continuo:** Es esencial mejorar la capacidad del detector para adaptarse a nuevos contextos utilizando diferentes tipologías de *deepfake*, evitando que pierda precisión al enfrentarse a

vídeos generados con técnicas o estilos distintos a los del entrenamiento original (*cross dataset*). A continuación, se enumera la aplicación técnica asociada al caso:

- Construir y mantener un repositorio interno con casos verificados, tanto reales como manipulados, organizados por tipo de deepfake (*face swap*, *reenactment*, *lip sync*, *difusores*) permite entrenar modelos más precisos y adaptables. Cada muestra se acompaña de metadatos relevantes como plataforma de origen, nivel de compresión, idioma y duración, lo que facilita el análisis contextual y la segmentación por escenarios. Este repositorio alimenta un ciclo de *active learning*, una técnica en la que el modelo identifica ejemplos ambiguos o inciertos durante la inferencia, que son revisados manualmente y reincorporados al entrenamiento. Esto mejora progresivamente la capacidad del sistema para detectar manipulaciones complejas y adaptarse a nuevos patrones.

- Reentrenar y validar periódicamente los modelos utilizando datasets de referencia (*DFDC*, *Celeb-DF*, *FaceForensics++*, *FakeAVCeleb*) permite mantener la eficacia frente a nuevas técnicas de manipulación. Estos conjuntos contienen ejemplos variados de deepfakes y sirven como base para evaluar el rendimiento del sistema. Además, incorporar *hard negatives*, vídeos auténticos que el modelo podría clasificar erróneamente como falsos, ayuda a reducir la diferencia entre los datos usados en el entrenamiento y los escenarios reales donde se aplicará el detector.

- Versionar modelos y medir su rendimiento por tipo de manipulación permite evaluar su precisión y estabilidad frente a diferentes técnicas de deepfake. Se utilizan métricas como *Area Under the Curve (AUC)* y *F1-score*, que indican la capacidad del modelo para distinguir entre vídeos reales y falsos, y el equilibrio entre precisión y exhaustividad. Además, se analiza la resistencia a la compresión (por ejemplo, niveles *CRF 23/40*), ya que los artefactos introducidos por la codificación del vídeo pueden dificultar la detección. Esta evaluación periódica permite ajustar los modelos y asegurar su fiabilidad en condiciones reales.

Posible caso de uso:

Tras un caso de deepfake con simulación de voz y labios en español, se añaden 300 vídeos etiquetados al repositorio interno, diferenciados por tipo de compresión (CRF 23/40) y metadatos relevantes. Se realiza un reentrenamiento del modelo multimodal, que combina señales de audio y vídeo para detectar inconsistencias espaciotemporales. El nuevo detector mejora un 8 % en AUC frente a vídeos reales no vistos en entrenamiento, reduciendo falsos positivos en entrornos de comprensión variable.

ANEXO V

Herramientas para la verificación de contenidos digitales

El presente anexo recopila las principales herramientas y recursos técnicos empleados en los procesos de verificación activa y trazabilidad de contenidos. Su finalidad es ofrecer un ciclo de verificación integral y reproducible que permita a cualquier persona o institución a identificar, contrastar y documentar contenidos sospechosos de desinformación.

CATEGORÍA	HERRAMIENTAS	USO PRINCIPAL	OBSERVACIONES
Fact-checking (ES)	Maldita.es Newtral EFE Verifica RTVE Verifica EUvsDisinfo	Buscar desmentidos, guías y casos.	Gratuitos; EUvsDisinfo enfocado a injerencia extranjera.
Extensiones navegador	NewsGuard InVID & WeVerify	Reputación de fuentes; análisis de vídeo/foto.	NewsGuard es de pago para algunas funciones; InVID es gratuito.
Búsqueda inversa	Google Images TinEye Yandex Images	Localizar origen/ uso previo de imágenes.	Complementar con metadatos y archivo web.
Buscador de verificaciones	Fact Check Explorer (Google)	Encontrar fact-checks ya publicados; formación.	Complementar con metadatos y archivo web
Vídeo y metadatos	Forensically	Extraer miniaturas/ fecha; análisis forense (ELA, clonación).	DataViewer útil para fotogramas; Forensically no siempre concluyente.
Metadatos EXIF	ExifTool EXIF.tools	Leer metadatos (cámara, fecha, GPS)	Metadatos pueden faltar o estar alterados.

Herramientas para la verificación de contenidos digitales

CATEGORÍA	HERRAMIENTAS	USO PRINCIPAL	OBSERVACIONES
Archivo web / trazabilidad	Wayback Machine archive.today	Ver versiones históricas; preservar enlaces.	Útil contra link rot y para comparar cambios.
Enlaces cortos / reputación URL	CheckShortURL Unshorten.it VirusTotal urlscan.io	Desplegar shorteners; detectar riesgos y redirecciones.	VirusTotal/urlscan muestran artefactos y dominios relacionados.
Geolocalización / contexto	Google Earth, Maps, Street View SunCalc Mapillary Nominatim (OSM)	Verificar lugar y hora (sombras, vías, rótulos).	Requiere criterio espacial y contraste visual.
Difusión y bots	Hoaxy Bot Sentinel	Mapear propagación; detectar cuentas inauténticas.	Cobertura y precisión sujetas a cambios de plataforma.
Prebunking / entrenamiento	Bad News (juego) EUvsDisinfo Quiz	Entrenar “olfato” y sesgos; alfabetización práctica.	Evidencia de utilidad para inoculación cognitiva.
Detección de deepfakes (pro)	Resemble Detect Reality Defender Sensity	Señalar manipulación mediante audio, vídeo e imagen.	Orientadas a equipos/empresas; precisión y acceso varían.

Nota metodológica: Ninguna herramienta es infalible. La verificación requiere triangular evidencias (búsqueda inversa, metadatos, archivo web y reputación de dominio) y documentar cada paso con trazabilidad (fecha, hora, fuente y responsable).

ANEXO VI

Ficha de verificación rápida y criterios de documentación

El presente anexo proporciona una plantilla estandarizada para documentar procesos de verificación activa y respuesta ante posibles casos de desinformación. Su uso garantiza la trazabilidad, reproducibilidad y custodia de evidencias, en coherencia con los protocolos **del Esquema Nacional de Seguridad (ENS)** y las prácticas de análisis del **DSN.**

CAMPO	DESCRIPCIÓN / DATOS A CUMPLIMENTAR
Identificador de caso	Código interno o número de registro asignado.
Claim / URL analizada	Texto del contenido o enlace principal a verificar.
Fecha y hora de detección	Registro temporal del momento de localización.
Fuente o dominio de origen	Medio, red social, autor o entidad emisora.
Tipo de contenido	□ Noticia □Imagen □ Vídeo □ Audio □ IA generativa □ Otro (especificar).
Tema o narrativa	Breve descripción del asunto o contexto (ej. sanitario, político, geopolítico, etc.).
Herramientas empleadas	Mencionar las utilizadas: búsqueda inversa, InVID, ExifTool, Wayback Machine, etc.
Evidencias recopiladas	Describir hallazgos técnicos o contextuales relevantes (con enlaces o capturas).
Clasificación final del contenido	□ Veraz □ Falso □ Engañoso □ Fuera de contexto □ No verificable.
Recomendación o acción propuesta	Ejemplo: publicación de corrección, alerta interna, escalado institucional.
Responsable de la verificación	Nombre, cargo o unidad responsable.
Fecha de cierre de análisis	Cuando se concluye la revisión.
Observaciones adicionales	Notas o incidencias relevantes del proceso.

Nota metodológica: *Esta ficha forma parte del protocolo de verificación y debe acompañar toda acción de verificación o respuesta. Se recomienda conservar los formularios en soporte digital seguro durante un periodo mínimo de dos años o según la normativa interna aplicable. Su uso contribuye a mantener una cadena de custodia documental coherente con las prácticas de análisis de inteligencia y comunicación institucional.*