SERGIO ARCE GARCÍA

LA ANATOMÍA DE LA DESINFORMACIÓN: REDES, BOTS Y ODIO

Universidad de Valladolid

LA ANATOMÍA DE LA DESINFORMACIÓN

Redes, Bots y Odio

Colección ***Comunicación***, nº 4

ARCE GARCÍA, Sergio

La anatomía de la desinformación : redes, bots y odio / Sergio Arce García. – Valladolid: Universidad de Valladolid, 2026

212 p.; 24 cm (Comunicación ; 4) ISBN 978-84-1320-417-8

1. Redes sociales (Internet) – Aspecto social 2. Internet – Aspecto social I. Arce García, Sergio, aut. II. Universidad de Valladolid, ed. III. Serie

004.78:316.77
316.77:004.78

SERGIO ARCE GARCÍA

LA ANATOMÍA DE LA DESINFORMACIÓN

Redes, Bots y Odio

EDICIONES Universidad de Valladolid

COLECCIÓN comunicación

En conformidad con la política editorial de Ediciones Universidad de Valladolid (http://www.publicaciones.uva.es), este libro ha superado una evaluación por pares de doble ciego realizada por revisores externos a la Universidad de Valladolid.

Logotipo de la colección: Pablo Berdón Prieto
Diseño de Cubierta: Pablo Berdón Prieto

ISBN: 978-84-1320-417-8
DOI: https://doi.org/10.24197/eduva.3136

Diseño: Ediciones Universidad de Valladolid

Ley de Brandolini o principio de asimetría de la estupidez:

“La cantidad de energía necesaria para refutar tonterías es un orden de magnitud mayor que la necesaria para producirlas”. (Brandolini, 2013)

Índice

Preámbulo

Hacía tiempo que rondaba por mi cabeza desarrollar el capítulo de libro que escribí en "El Discurso de odio como arma política", que editó la profesora Virginia Martín Jiménez de la Universidad de Valladolid. Tanto ella, como otra mucha gente, me animaron a profundizar más en qué es y cómo funciona la actual industria de la desinformación y odio en internet, especialmente en redes sociales, pero que también alcanza a videojuegos, webs, blogs y un largo etc. Este es el fruto de todo ello.

Este libro está estructurado en dos grandes partes, independientes, pero totalmente complementarias. Una no puede entenderse completamente sin la otra. En una primera parte se expone la historia de la desinformación y la propaganda, especialmente en la época de internet. Se incluyen igualmente pasajes anteriores históricos que ayuden a entender y ubicar el porqué de cada cosa. Este recorrido histórico pretende ayudar a entender qué está pasando en el mundo y cómo se libran batallas a nivel geopolítico, empresarial y social, donde ya existen centenares de empresas que ofrecen sus servicios para hundir al adversario o polarizar sociedades enteras, a un precio cada vez más bajo.

La segunda parte tiene un enfoque distinto al de muchas otras obras, ya que en este caso la intención es mostrar cómo actúan, desinforman y provocan odio, para entender y comprender cómo trabajan. En este caso, más que enseñar a verificar si algo que nos llega por el teléfono móvil es real o no, aquí mostraremos cómo lo hacen, para aprender a identificar sus métodos y formas.

Obviamente no se puede reflejar todo, es imposible por extensión y porque es algo a escala mundial, pero al menos exponer una fotografía que muestre un mundo tan desconocido para la mayoría. Sé, además analizando una industria cuya misión es odiar y polarizar, que este libro puede ser interpretado de muchas maneras según cada persona que lo lea y podría tener diferentes reacciones. La intención es la de mostrar una serie de hechos y métodos, con fuentes contrastables que, ojalá, puedan ayudar a combatir este gran problema de la desinformación y el odio en la sociedad mundial.

Conviene resaltar que todo lo expuesto se basa en informes oficiales, prensa contrastada y literatura académica relevante, así como investigaciones académicas en las que participé, pudiendo en algunos casos ser fuente de controversia o interpretación debido a los temas que trata. El objetivo no es más que intentar hacer comprender la arquitectura de las campañas algorítmicas de manipulación o influencia, con el fin de defender la democracia y los derechos humanos, y que no se pretende estigmatizar ideologías, religiones, pueblos ni minorías. No se juzga en ningún momento colectivos ni poblaciones enteras, sino que cuando se nombran es porque pueden ser instrumentalizados por actores de desinformación.

Los nombres de usuario y mensajes reproducidos corresponden a publicaciones realizadas en plataformas de acceso público, cuyos titulares hicieron manifiestamente públicos. Su inclusión se ampara en los artículos 9.2.e) y 9.2.j) del Reglamento (UE) 2016/679 (RGPD), limitándose a los ejemplos estrictamente necesarios para documentar los hallazgos del análisis con fines de investigación científica.

Espero les guste y, al menos, les haga reflexionar.

1. ¿Cómo hemos llegado hasta aquí?

1.1. LA DESINFORMACIÓN NO ES ALGO DE AHORA

El complejo de templos de Abu Simbel, ubicado al sur del actual Egipto, cerca de la frontera con Sudán, es uno de los destinos más visitados del país para explorar su rica historia faraónica. Mandado construir por Ramsés II en el siglo XII a.C., este faraón es uno de los más célebres y con mayor influencia en la civilización egipcia. Abu Simbel, que consta de un templo dedicado a Ramsés y otro a su esposa Nefertari, fue tallado en la roca del desierto con el propósito de anunciar a los pueblos de Nubia y del Alto Nilo que, al seguir el río, estaban ingresando en las tierras egipcias. El tamaño y el arte de estos templos reflejan el poder de Ramsés, destacando sus victorias en la batalla de Qadesh contra los hititas. Las paredes del templo están adornadas con grabados que muestran a numerosos hititas prisioneros y ejecutados, glorificando a un Ramsés de tamaño y fuerza sobrehumana, acompañado por los principales dioses egipcios. Sin embargo, la gran victoria que se celebra en estos relieves nunca ocurrió realmente tal cual...

La batalla de Qadesh es la primera de la que se conoce su desarrollo paso a paso y movimientos estratégicos, y aunque ambos ejércitos sufrieron grandes pérdidas, según los historiadores terminó sin un vencedor. Se sabe que, al regresar Ramsés a Egipto desde el campo de batalla, situado entre la actual Siria y Líbano, el ejército fue abucheado por su pueblo debido a las pérdidas sufridas. A pesar de esto, la batalla fue transformada en una gran victoria, convirtiéndose en un instrumento de propaganda que alteró la percepción de otros pueblos e incluso del propio con el tiempo. Hoy en día, Abu Simbel sigue presentando a un Ramsés II majestuoso y victorioso, haciendo sentir pequeños a todos los visitantes de estos templos. Este ejemplo ilustra cómo la manipulación de la realidad no es un fenómeno reciente, sino que a lo largo de la historia existen numerosos casos que demuestran cómo se puede alterar cómo se perciben los acontecimientos.

Igualmente desde la antigüedad, pero procedente desde la otra parte del mundo, nos llega el libro "El Arte de la Guerra", escrito entre los siglos VI y V a.C., obra de Sun Tzu, y que aún hoy día se sigue estudiando en muchas escuelas

militares mundiales. Su autor teoriza sobre la manipulación de la información, subrayando que la excelencia suprema consistía en "romper la resistencia del enemigo sin luchar". Su obra sostiene que la guerra siempre comienza en la mente, y si se controla la percepción, la moral y la información, se ganará sin enfrentamiento directo.

Por tanto, la influencia y manipulación de la población, e incluso de países enteros, tienen múltiples ejemplos a lo largo de la historia, no podemos decir que sea algo reciente. Pero que no sea reciente, que lleve siglos haciéndose, no significa que no fuera evolucionando y aprovechando las transformaciones de las sociedades y sus nuevas tecnologías. Así, la llegada de los medios de comunicación y las sociedades modernas cambiaron la perspectiva de esta manipulación en el siglo XIX, donde tenemos un buen ejemplo a finales de ese siglo cuando los periódicos de Estados Unidos alentaron la guerra en Cuba contra España (como el *New York World* de Pulitzer) (Campbell, 2001), llegando incluso a grabarse en unos estudios cinematográficos en Brooklyn, Nueva York, la explosión de una maqueta que supuestamente representaba el fuerte de La Habana, así como el arriado de la bandera española e izado de la estadounidense para ser emitidos en los primeros cines de la historia (Rodrixoc, 2019).

1.1.1. Edward Bernays, el arquitecto de la opinión pública moderna

La llegada del siglo XX trajo mayores y profundas transformaciones sociales y políticas, surgiendo a su vez nuevas teorías y enfoques para comprender el funcionamiento de la cada vez más importante opinión pública. Y para ello debemos introducir a Edward Louis Bernays, doble sobrino de Sigmund Freud (su madre era hermana del célebre psicoanalista y su padre, hermano de la esposa de Freud), que se consolidó como fundador de las relaciones públicas y tuvo gran influencia en numerosas campañas en muchos años posteriores, dejando una huella duradera en la manera en que gobiernos de muchos países y empresas abordaron la persuasión y la gestión de la opinión pública. Figura controvertida e influyente, Bernays desempeñó un papel decisivo tanto en la política como en el desarrollo del marketing empresarial, a pesar de seguir siendo en gran medida desconocido para el público general a día de hoy.

Nacido en Viena en 1891 y emigrado a Estados Unidos en la infancia, Bernays desarrolló a lo largo de su extensa vida (falleció en 1995 a los 103 años) una teoría y práctica revolucionaria que transformaría las relaciones entre gobiernos, corporaciones y ciudadanía a través de la "ingeniería del consentimiento",

concepto que sitúa una pregunta inquietante sobre los límites entre persuasión legítima y la manipulación sistemática dentro de una sociedad democrática.

Sus primeras aproximaciones a este campo comenzaron durante la Primera Guerra Mundial, cuando participó en el Comité de Información Pública de Woodrow Wilson, organismo que constituyó la primera maquinaria estatal de propaganda masiva en la historia estadounidense y que ayudó a convencer al pueblo a entrar en dicha guerra. Al terminar el conflicto, Bernays estableció en 1919 su propia consultoría, distanciándose deliberadamente del término "propaganda" para acuñarse a sí mismo como "consejero en relaciones públicas". Esta designación reflejaba su ambición de presentar su labor como disciplina científica fundamentada en psicología social, psicoanálisis freudiano y estudios sobre comportamiento colectivo. Bernays sintetizó una arquitectura conceptual profundamente pesimista sobre la racionalidad de las masas, tomando de su tío Freud la premisa de que los seres humanos están gobernados por fuerzas inconscientes, deseos reprimidos y motivaciones ocultas que escapan al control racional. De otros autores como Le Bon y Trotter, adoptó la tesis de que en multitudes los individuos pierden el pensamiento crítico y se vuelven extraordinariamente sugestionables, siguiendo únicamente sus emociones. De Walter Lippmann y su obra *Public Opinion* (Lippmann, 1922), internalizó la idea de que ciudadanos carentes de recursos cognitivos requerían una "clase especializada" de expertos que modelara su opinión pública, ya que su visión del mundo se forma a partir de imágenes, estereotipos y narrativas simplificadas. De esta manera, Bernays declaró en su obra Propaganda (1928) que "La manipulación consciente e inteligente de los hábitos y opiniones organizadas de las masas es un elemento de importancia en la sociedad democrática. Quienes manipulan este mecanismo oculto constituyen un gobierno invisible que detenta el verdadero poder".

Su práctica profesional produjo campañas muy destacadas, como las "Antorchas de Libertad" de 1929, donde vinculó el poder fumar de las mujeres con su derecho al voto, a través de la organización de un desfile altamente publicitado en la Quinta Avenida de Nueva York donde mujeres destacadas encendían cigarrillos de la marca Lucky Strike como símbolo de libertad. Simultáneamente, en un trabajo para la *Beech-Nut Packing Company*, Bernays convocó a cinco mil médicos para debatir si los desayunos abundantes eran saludables, distribuyendo luego sus respuestas afirmativas como titulares que proclamaban que la medicina recomendaba huevos con tocino, estableciendo una tradición culinaria estadounidense que pervive a día de hoy, como es el desayunar con *bacon*. La intervención más polémica y oscura de Bernays tuvo lugar en Guatemala, cuando la *United Fruit Company* lo contrató en 1951 para construir una

imagen pública negativa sobre el gobierno legítimo de Jacobo Árbenz, presentándolo como una amenaza comunista. Bernays diseñó una sofisticada campaña internacional de desinformación, ofreciendo desde difusión de artículos bajo seudónimo a la organización de viajes guiados para periodistas con información manipulada, o la elaboración de informes para autoridades estadounidenses que denunciaban un peligro soviético inexistente. Toda esta estrategia sentó las bases en la opinión pública y política de Estados Unidos para que el gobierno de Eisenhower respaldara el golpe de la CIA sucedido en 1954. Aquella operación marcó el inicio de una larga y brutal guerra civil en Guatemala que costaría la vida a unas 200.000 personas y dejaría una herida profunda por su violencia (Schlesinger, & Kinzer, 1982).

Según Bernays, incluso la causa social más noble está condenada al fracaso si no logra ganarse el apoyo de la opinión pública. En otras palabras, el éxito y la legitimidad de cualquier iniciativa dependen de que las acciones de individuos o instituciones estén en sintonía con las expectativas, necesidades y emociones colectivas de la sociedad. Aunque en teoría cada persona debería elegir los productos más eficientes y baratos que ofrece el mercado, en la práctica esa racionalidad es inviable. Por eso, la economía real y cotidiana está mediatizada por atajos mentales, hábitos de grupo y la confianza depositada en recomendaciones o figuras de referencia. Bernays insistía en que "la mente del grupo no piensa, en sentido estricto: obedece impulsos, hábitos y emociones". Y llegados a contextos de incertidumbre, la decisión individual se pliega al ejemplo de un líder o al sentir colectivo, fenómeno que la psicología de masas ha documentado ampliamente. Así se explicaría, por ejemplo, el auge o la caída de la reputación de un hotel, la estampida en un banco, el pánico bursátil o el nacimiento de fenómenos de consumo masivo como bestsellers y superproducciones.

Siguiendo las ideas de Bernays, los métodos de persuasión y propaganda funcionan porque aprovechan la tendencia natural de las personas a adoptar las ideas, valores y aspiraciones de los grupos con los que se identifican. Así, cuando llega el momento de decidir, ya sea en política, negocios o consumo, la mayoría no actúa tras un análisis totalmente racional, sino que se guía, casi siempre, por la conexión emocional y el sentido de pertenencia que mantiene con el entorno o colectivo que considera propio (Bernays, 1955).

Uno de los aspectos más inquietantes del legado de Bernays es que su obra fue adoptada como referencia por Joseph Goebbels, el ministro alemán de propaganda nazi, quien empleó sus teorías y métodos para construir campañas que justificaron las persecuciones y atrocidades del régimen en Alemania (y al que se le atribuye popularmente la famosa frase "Una mentira repetida mil veces se

convierte en verdad"). Pese a su origen judío y al desconcierto que esto le produjo en un inicio, Bernays consideraba sus técnicas como herramientas neutrales, separando siempre la capacidad de persuadir de la reflexión ética sobre su posible utilización. Esa distancia entre la eficacia y la moral la mantuvo toda su vida, pues nunca expresó un arrepentimiento por las consecuencias humanas de sus campañas. Así, la vida y obra de Bernays encarnan la gran paradoja de la modernidad, pues mientras que el ideal democrático se apoya en la autonomía del ciudadano informado, la realidad revelaría cómo unas élites pueden modelar de manera invisible la opinión pública mediante la manipulación emocional. Este problema, en la era de la desinformación viral y los algoritmos de microsegmentación, se vuelve más vigente que nunca (Tye, 1998).

1.1.2. La teoría de la aguja hipodérmica y su evolución

Al mismo tiempo que Bernays, la teoría de la bala mágica o de la aguja hipodérmica surge en las primeras décadas del siglo XX como una de las explicaciones más influyentes sobre el poder de los medios de comunicación en la sociedad de masas. Formulada a partir de las experiencias de propaganda en la Primera Guerra Mundial y los estudios pioneros de Harold Lasswell, esta teoría planteaba que los mensajes emitidos por los medios actúan como proyectiles que penetran directamente en la mente de los individuos, produciendo respuestas uniformes, inmediatas y potentes. En este modelo, la audiencia no es crítica ni reflexiva, sino que se ve sometida, casi de forma mecánica, por la influencia de la manipulación mediática.

Esta concepción se apoya en la idea (compartida por Bernays y muchos pensadores contemporáneos suyos) de que las masas son vulnerables, poco racionales e increíblemente susceptibles a las emociones, los símbolos y los líderes carismáticos. Para los defensores de la teoría de la aguja hipodérmica, el individuo aislado en una multitud urbana carece de los filtros sociales tradicionales y se convierte en un receptor pasivo, dispuesto a aceptar como verdaderas las narrativas propuestas por los grandes medios de prensa, radio o cine.

Uno de los ejemplos más ilustrativos de la aplicación de esta teoría citados por los manuales de comunicación, es la emisión radiofónica de Orson Welles de la obra "La Guerra de los Mundos". Fue emitida en 1938 por la cadena *Columbia Broadcasting System* (CBS) y basada en la novela homónima de H.G. Wells, donde la transmisión simulaba un noticiero en tiempo real que relataba la invasión de Estados Unidos por extraterrestres. Pese a las advertencias iniciales y la naturaleza ficticia del contenido, muchos oyentes entraron en

pánico, creyendo más de un millón de personas que el ataque era real. Durante horas cundió el miedo en varias ciudades, la gente huyó de sus casas, colapsaron líneas telefónicas y los hospitales recibieron múltiples avisos de crisis nerviosas. El episodio ilustró (quizá de forma algo exagerada por la posterior mitificación periodística) el alcance devastador que podía tener un mensaje persuasivo emitido desde un medio de masas, reforzando la creencia de los efectos inmediatos e irresistibles de la comunicación.

Aunque las investigaciones posteriores, impulsadas por autores como Paul Lazarsfeld y la Escuela de Columbia, matizaron el supuesto poder total de los medios introduciendo nociones como los líderes de opinión y el flujo de comunicación, la imagen de la aguja hipodérmica o de la bala mágica perdura como una advertencia sobre la fuerza que pueden alcanzar los mensajes mediáticos en contextos de credulidad social y ansiedad colectiva. El caso de Orson Welles sigue fascinando a historiadores y teóricos de la comunicación como ejemplo de la sugestión poderosa (y a veces peligrosa) que esconde toda tecnología de masas.

Tras la vigencia de esta teoría, el campo de la comunicación de masas vivió una transformación radical tras la Segunda Guerra Mundial. Los investigadores dejaron atrás la idea de una audiencia homogénea y pasiva, para dar paso a modelos donde la segmentación o clasificación en grupos de la sociedad se volvieron protagonistas. Estas nuevas perspectivas demostraron empíricamente que no todos los receptores reaccionan igual ante los mensajes de los medios, ya que su influencia está condicionada por factores individuales, sociales y contextuales (Wolf, 1987).

Esta concepción abrió la puerta a nuevas formas de clasificación de las audiencias, pues ya no se trataba sólo de emitir mensajes masivos, sino de reconocer nichos definidos por edad, género, ubicación geográfica, intereses o hábitos de consumo. Así, el estudio de la segmentación de audiencias permitió a publicistas, políticos y programadores optimizar sus estrategias, haciendo que el mensaje fuera más específico, personalizado y eficaz para cada subgrupo. Con el avance de las tecnologías digitales y el análisis de datos, esta tendencia se profundizó, dando lugar a sistemas de microsegmentación que constituyen hoy la base de la comunicación política y comercial contemporánea.

La teoría que en la actualidad tiene más difusión es la *agenda-setting*, que sostiene que los medios de comunicación no establecen qué se debe pensar, pero sí sobre qué se debe pensar, hablar y de qué debemos preocuparnos. Al destacar unos temas sobre otros, su orden de importancia y cuáles se ignoran, influyen de manera decisiva en aquello que el público percibe como importante o digno de

atención. Complementando a esto, la teoría del *framing* (encuadre o marco) explica que los medios no sólo seleccionan los temas, sino también la manera en que los presentan, pues enfatizan ciertos aspectos, interpretaciones o valores, orientando así cómo las audiencias deben entender y debatir la realidad.

1.1.3. Las oleadas de la desinformación moderna

Si pretendemos llevar todo lo descrito hasta ahora al mundo de la desinformación organizada, primeramente deberemos establecer el proceso histórico de evolución de las campañas de desinformación. Para ello deberemos pararnos ante la obra de Thomas Rid, que divide en cuatro oleadas a la desinformación moderna, revelando una historia de sombras en la que se entrelazan tecnología, ideología y poder. Este autor, profesor de Estudios Estratégicos en la *Johns Hopkins University* en Estados Unidos, propone en su libro *Active Measures: The Secret History of Disinformation and Political Warfare* (publicado en español como "Desinformación y guerra política: Historia de un siglo de falsificaciones y engaños") (Rid, 2020) una genealogía de la desinformación estructurada en cuatro fases, separadas aproximadamente por una generación. En su obra Rid establece que la mentira no surge del caos, sino del método, y no brota de la improvisación política, sino de la planificación burocrática de los servicios de inteligencia.

La primera oleada la establece en el turbulento periodo de entreguerras mundiales, cuando la radio transformó el periodismo y amplificó la voz de los estados. En ese tiempo de crisis económica y convulsión ideológica, la creatividad de la desinformación se canalizó hacia operaciones de apariencia conspirativa. Este contexto llevó a la creación del primer organismo dedicado a estos temas a finales de 1917, cuyo objetivo era realizar propaganda de influencia principalmente frente a los rusos zaristas que habían abandonado el país tras la revolución: la Cheká (Всероссийская Чрезвыча́йная Комиссия - ВЧК), que significa en su versión extendida "Comisión Extraordinaria Panrusa para la Lucha contra la Contrarrevolución, la Especulación y los Delitos en el Cargo" (Russian Presidential Library, n.d.). Este organismo se estableció para gestionar la propaganda y controlar la narrativa en un momento de gran agitación política, con la intención de eliminar cualquier acto que fuera contrarrevolucionario o se desviara de la línea oficial. Una de las actividades más notorias de la Cheká fue la Operación Confianza (*Operatsiya Trest*), emprendida entre 1921 y 1927, cuando una falsa organización monárquica, la *Monarchist Union of Central Russia*, tendió una ingeniosa trampa a exiliados antibolcheviques y a servicios occiden-

tales. Además de provocar la muerte del agente británico Sidney Reilly, la operación permitió neutralizar a opositores, dividir a la emigración rusa zarista y demostrar ante el mundo la capacidad soviética para convertir el engaño en arma de Estado. Aquella primera generación de falsificaciones mostró que la desinformación era, en el fondo, un arma barata que servía tanto para defenderse como para atacar.

Su estructura fue copiada en otros países, como en la España de la Guerra Civil, donde se denominaba como "checas" a los lugares de represión política mediante interrogatorios, torturas, ejecuciones y juicios sumarísimos vinculados a diversos grupos en la zona republicana (donde solo en Madrid 1.000 chequistas asesinarían a unas 3.000 personas) (Instituto CEU de Estudios Históricos, 2012). Igualmente llegó a situarse la existencia de "checas azules" en el bando sublevado franquista en Sevilla (García Márquez, 2010).

La preocupación por este tipo de campañas empezó a extenderse y llegó incluso a ser debatido en el Parlamento británico el 3 de marzo de 1925 bajo el título "*Subversive Propaganda (Great Britain and the Empire)*" (House of Commons, 1925), que abordaba la preocupación por la expansión de la propaganda subversiva a lo largo de Gran Bretaña y sus colonias. Varios parlamentarios manifestaron su inquietud por la difusión de ideas consideradas peligrosas para el orden político y social establecido, señalando especialmente la influencia de movimientos revolucionarios extranjeros y el uso estratégico de medios impresos, panfletos y prensa para llegar a sectores populares. Se pidió una mayor vigilancia y políticas más firmes para contrarrestar esas campañas, apelando a la defensa de la estabilidad nacional y la protección de los valores imperiales frente a lo que se percibía como una amenaza coordinada e internacional (Briant, & Jones, 2025).

Tras la Segunda Guerra Mundial la desinformación entró en su segunda oleada, según Rid, en la era de la profesionalización. Durante la Guerra Fría, los servicios de inteligencia (especialmente el KGB soviético y la CIA estadounidense) dominaron el arte del encubrimiento informativo. Las "medidas activas" (forma en cómo los servicios rusos denominaban a estas campañas) se institucionalizaron y se insertaron en un tablero global donde cada falsa historia servía a un propósito geopolítico.

Fue entonces cuando los especialistas del engaño mejoraron sus técnicas, como mezclar verdades y falsedades, para minar la confianza y fomentar divisiones internas. Revistas creadas como tapaderas, panfletos falsos y periódicos clandestinos se convirtieron en herramientas de esta nueva diplomacia encubierta. Rid recuerda, por ejemplo, los panfletos falsamente atribuidos al Ku Klux Klan y distribuidos por el KGB entre delegaciones africanas y asiáticas en la ONU

en 1960, una maniobra destinada a desacreditar a Estados Unidos frente al mundo del postcolonialismo. En este contexto, se rescata igualmente la historia de un excapitán de submarinos nazi que, instalado en la Alemania Occidental tras la derrota, recibió apoyo de los servicios de inteligencia occidentales para fundar revistas dirigidas a públicos específicos, como obreros, intelectuales y jóvenes de la Alemania Oriental. Aquellas publicaciones, bajo apariencia independiente, formaban parte de una sofisticada estrategia de manipulación en la que no se exponía propaganda de forma general, sino segmentada en función del tipo de población. Su objetivo era erosionar la cohesión social alemana oriental y orientar la opinión pública hacia intereses externos. Era la guerra fría de las palabras, donde se evoluciona hacia la segmentación de objetivos, ya no se ofrecía publicidad hacia un público general homogéneo.

En esta época del engaño cabe recordar también los antecedentes más artísticos y experimentales de la propaganda psicológica, como los que rescata Peter Pomerantsev en su libro *How to Win an Information War: The Propagandist Who Outwitted Hitler* (Pomerantsev, 2024). En este trabajo el autor muestra la figura de Sefton Delmer, periodista y agente británico que durante la Segunda Guerra Mundial dirigió la emisora de radio denominada *Gustav Siegfried Eins* (abreviado GS1) y otras estaciones clandestinas, estableciendo la denominada "propaganda negra" (o *dark o black PR*, que dará posteriormente a un tipo de campañas negativas contra gobiernos y empresas (Rodríguez Fernández, 2021, Ennis, 2023)). Se fingía ser una transmisión clandestina desde dentro de la Alemania nazi, pero que en realidad estaba producida por los británicos con el objetivo de socavar la moral y la cohesión del régimen. La voz principal era la de Peter Seckelmann, un refugiado berlinés que interpretaba a "*Der Chef*", un supuesto oficial patriota prusiano, crítico feroz de los burócratas del Partido Nazi a quienes describía como corruptos y decadentes, en contraste con la heroicidad de los soldados alemanes en el frente.

La emisora difundía burlas, sátiras y rumores sobre la corrupción del mando alemán junto a mensajes oficiales del régimen nazi, mezclando información con engaño. El equipo de colaboradores no solo se encargaba de interpretar personajes y escribir guiones, sino que además recreaba con gran autenticidad la jerga y los temas internos del nazismo para que la propaganda resultara convincente. Buena parte del equipo de Delmer estaba integrado por refugiados judíos alemanes, así como artistas y guionistas del cabaret berlinés que se habían exiliado en el Reino Unido para huir de la persecución nazi, y que colaboraban precisamente para subvertir la narrativa del régimen desde dentro. Entre los colaboradores de Delmer se encontraba Ian Fleming, el futuro creador del agente 007 James Bond,

quien contribuyó con ideas de espionaje narrativo y manipulación psicológica. El éxito fue extraordinario, ya que muchos soldados alemanes creyeron escuchar a un camarada desilusionado, no a un enemigo, llegando Goebbels a reconocer tras la guerra que aquellas emisiones dañaron la moral nazi más que muchas bombas. Era atacar a la propaganda con sus propias armas.

La tercera oleada según Rid, que se desarrolló entre los años setenta y ochenta, llevó la desinformación a su máxima sofisticación logística e intelectual. Las operaciones contaron con ingentes recursos y una coordinación sin precedentes. Una de las más célebres fue la Operación Denver, mediante la cual el KGB difundió la idea de que el virus del SIDA era un arma biológica creada por Estados Unidos en el laboratorio de Fort Detrick. Aquella narrativa conspirativa alcanzó un eco amplio, en especial entre comunidades históricamente desconfiadas del poder y conspiranoicas, y que según Rid fue "la medida activa más exitosa de los años ochenta". Las campañas de esa década se multiplicaron hasta cifras de decenas de miles, con una precisión y una extensión que auguraban lo que vendría en la era digital.

Esa cuarta oleada llegó en torno a 2010, bajo el signo de las redes sociales y el auge de la cultura digital. Internet introdujo nuevas velocidades y nuevas máscaras, donde la combinación de hackeo, filtración y viralidad convirtió la mentira en un producto de consumo masivo. Rid llegó incluso a advertir de manera pública meses antes de los comicios de Estados Unidos en 2016, que ganaría Donald Trump, que los servicios rusos estarían "planificando cuidadosamente una campaña política de alto riesgo" en dichas elecciones.

El resultado de esta última ola fue un paisaje comunicativo donde las operaciones se volvieron más baratas, rápidas y difíciles de rastrear. A lo largo de este siglo de operaciones ocultas, la desinformación ha demostrado ser una fuerza paciente, que erosiona la confianza poco a poco. Rid subrayó que las "medidas activas" no eran mentiras improvisadas, sino productos sistemáticos de grandes burocracias entrenadas en el arte del engaño. Su meta última no es convencer, sino dividir, sembrar sospecha entre aliados, profundizar grietas sociales, corroer la autoridad de las instituciones y minar la idea misma de verdad.

1.1.4. Yuri Bezmenov y la subversión ideológica

Para complementar cómo se trabajaba y el grado de sofisticación alcanzado a lo largo de las oleadas segunda y tercera establecidas por Rid, hay un personaje que ilustra muy bien dichos métodos. Yuri Bezmenov fue un disidente soviético, antiguo miembro de la agencia de noticias rusa RIA Novosti y experto en

propaganda y desinformación del KGB que desertó en los años 70 y, desde entonces, dedicó su vida a advertir sobre los métodos de subversión ideológica y manipulación social empleados por la Unión Soviética en Occidente. En sus entrevistas y conferencias de la década de 1980 (que pueden verse en Youtube), Bezmenov expuso una teoría sobre el verdadero alcance de la guerra fría, pues sostenía que el espionaje tradicional era solo una fracción menor de la actividad soviética, y que el verdadero esfuerzo del KGB se concentraba en la "subversión ideológica". Según este disidente, este proceso de subversión no es un acto oculto ni repentino, sino una transformación lenta y a menudo visible cuyo objetivo es reconfigurar poco a poco la percepción de la realidad tanto de una nación como de sus ciudadanos, de manera que lleguen a perder sus valores, sentido crítico y capacidad de autodefensa social. La subversión se realizaría en cuatro etapas:

1. Desmoralización: consiste en un trabajo sostenido durante quince a veinte años, el tiempo suficiente para transformar los valores de toda una generación a través de la educación, los medios y el discurso público. Su objetivo principal es que la sociedad pierda sus referencias tradicionales, confianza en sus instituciones y sentido crítico, reemplazando estos elementos por una mezcla de apatía, relativismo y confusión moral. Según Bezmenov, una vez desmoralizada, la población se vuelve incapaz de distinguir la información verdadera de la falsa y los hechos de las opiniones, quedando indefensa ante la manipulación.
2. Desestabilización: fase más breve, que se centra en debilitar las estructuras fundamentales como la economía, las fuerzas armadas y las alianzas diplomáticas. Aquí, los agentes de influencia buscan socavar la confianza en el sistema, infiltrando organizaciones sociales, ONGs y medios prestigiosos, y fomentando conflictos internos que paralicen la toma de decisiones y siembren el caos en el liderazgo.
3. Crisis: suele desarrollarse en cuestión de semanas y busca provocar un colapso abrupto del orden institucional, generando condiciones para una "solución" que a menudo implica el establecimiento de un nuevo orden autoritario o el sometimiento a intereses externos.
4. Normalización: periodo en que el régimen, tras haber tomado el poder, presenta el nuevo estado de cosas como inevitable, legítimo y seguro, aunque implique la represión de opositores y la imposición de controles estrictos sobre la sociedad.

Bezmenov señaló que esta forma de "guerra psicológica" podía durar décadas y que la mayoría de las actividades del KGB en Occidente estaban orientadas a influir en ámbitos como la educación, los medios de comunicación, las élites políticas y culturales, más que en espionaje tradicional. Igualmente también advirtió contra el apoyo pasivo o involuntario que, según él, el mundo occidental brindaba a los regímenes totalitarios a través de acuerdos económicos, tecnológicos y culturales, señalando una "paradoja histórica" en la que el capitalismo financiaba la maquinaria de su propio potencial destructor.

Un aspecto particular que también señaló fue que no trataban de convencer ni emplear a los comunistas idealistas locales de otros países ("idiotas útiles", según la jerga del KGB), pues estos resultaban demasiado volátiles, ya que una vez que comprobaban las diferencias entre la retórica comunista y la realidad del poder, podían volverse los críticos más peligrosos del sistema. Por ello, Bezmenov señalaba que el KGB buscaba personas influyentes, carismáticas o con ambiciones personales, pero desprovistas de verdadera convicción ideológica. Tanteaban académicos dispuestos a firmar manifiestos, periodistas cínicos, artistas deseosos de reconocimiento, o políticos pragmáticos, pues eran mucho más controlables y útiles a largo plazo en la tarea de penetración ideológica y subversión paulatina (Quiñones de la Iglesia, 2021).

Su visión (que fue muy influyente en círculos anticomunistas de la guerra fría y en la cultura política contemporánea) insistía en la responsabilidad del individuo y la sociedad civil para reconocer y resistir los mecanismos de propaganda, así como en la necesidad de mantener la educación cívica y los valores democráticos frente a la manipulación ideológica y el avance de regímenes autoritarios.

1.1.5. De 1990 al 2000

Diversos hechos comienzan a cambiar la situación en el mundo que se conocía hasta entonces, pero algunos determinan mucho en la industria que nos ocupa. Partimos en 1990 con Newt Gingrich, un político norteamericano que llegó a ser presidente de la cámara de representantes entre 1995-1999 y candidato a la presidencia del Partido Republicano en 2012, que estableció un listado de palabras en el famoso memorando "*Language: A key mechanism of control*" para definir a sus oponentes (The American Leader, 1990). Dichas palabras eran claras, fáciles de entender y de introducir en los discursos contra oponentes políticos (radical, traidor, sensacionalista, intolerante, colapsar, corrupto, mentir, etc.)(Pinkola Estés, 2012), creando polarización frente a palabras optimistas para referirse a

uno mismo. Sus métodos y formas a la hora de comunicar tuvieron mucha trascendencia en la política comunicativa del partido hasta hoy día.

Lejos de allí las nuevas tecnologías, y especialmente un internet muy joven iba a tener, por primera vez, influencia sobre hechos históricos. Durante el golpe de Estado en Rusia del 19 de agosto de 1991, cuando el KGB y altos funcionarios del Estado intentaron destituir a Gorbachov (presidente de la Unión Soviética) y tomar el control, los programadores de Relcom (el internet ruso, creado a finales de los 1980 y conectado al internet mundial en 1990 a través de Finlandia) jugaron un papel clave en la difusión de información y la coordinación de la resistencia (Soldatov, & Borogan, 2015). Con los medios censurados y Yeltsin (presidente de Rusia entre 1991 a 1999) atrincherado en la Casa Blanca rusa, Relcom permitió el envío masivo de comunicados y noticias tanto dentro como fuera del país, incluyendo proclamaciones de Yeltsin y mensajes de ánimo y apoyo internacional.

Relcom funcionó como canal alternativo frente a la censura del momento, activando incluso una iniciativa llamada "Regime N1", que pedía a los usuarios informar desde sus ventanas sobre lo que observaban, recopilando datos sobre la situación en diversas ciudades. En tres días, transmitieron más de 46.000 noticias que fueron fundamentales para coordinar la resistencia y evidenciar que los golpistas apenas controlaban las ciudades de Moscú y Leningrado. La estructura horizontal y descentralizada de la red, así como la decisión de mantener el canal abierto pese a los riesgos, permitieron que la información independiente circulara por todo el país y al extranjero. Era una innovación revolucionaria en una sociedad históricamente controlada y censurada. El fracaso del golpe fue atribuido, en parte, a esta nueva capacidad para difundir información y organizarse rápidamente fuera del aparato oficial del Estado, mostrando por primera vez el poder social de internet como herramienta de resistencia política.

Volviendo a Estados Unidos, en 1996, el ejército formaliza el concepto de "Dominación Informativa" (*Information Dominance*) en su doctrina militar, definiendo la superioridad en el control, acceso y manipulación de información como un elemento crítico en cualquier conflicto moderno (Colon, 2025). Se trató de considerar los datos y la comunicación como recursos tan estratégicos como las armas físicas, permitiendo influir, perturbar y, en última instancia, controlar el entorno operativo del adversario, tanto a nivel militar como civil.

Ese mismo año, China abre el acceso público a Internet, pero casi de inmediato comienza a construir mecanismos para regular y censurar la circulación de información, sentando las bases de su sistema de censura conocido como

"Gran Muralla Cortafuegos" o como *Golden Shield*. Este sistema fue implementado en 1998 y evolucionó hacia una infraestructura nacional capaz de bloquear, filtrar y vigilar contenidos considerados potencialmente peligrosos.

En el año 1997 el coronel ruso Sergei Komov publicó en la revista *Voennaia mysl'* (Mentalidad militar) los 11 principios fundamentales que sustentan el control reflexivo, como medio de influencia en la toma de decisiones del adversario (Thomas, 2004): (1) distraer al enemigo creando amenazas reales o ficticias, (2) sobrecargarlos de información, (3) paralizar infundiendo miedo, (4) agotamiento forzando que haga operaciones inútiles, (5) disminuir su vigilancia por repetición de amenazas ficticias, (6) dividir internamente mediante conflictos, (7) engañar sugiriendo que lo que se ve no es real, (8) intimidar mediante acciones de peso que crean la sensación de inmensurable superioridad, (9) provocar a realizar acciones que beneficien al atacante, (10) sugerir e insinuar mediante propaganda y desinformación, y (11) desacreditar a los gobiernos hostiles ante su público. Este esquema marcaría la doctrina posterior rusa, que vendría de una evolución desde los años 60 a una fase cada vez más psicosocial, que distintos investigadores declaran mucho más sofisticada y orientada a controlar que a influir. Mientras, la defensa occidental es mucho más reactiva, parcial, y que reacciona tarde y superficialmente (Institute for the Study of War, 2015; Institute for the Study of War, 2025; Pontijas, 2020).

En 1998, la percepción de la información como "arma estratégica" tomó fuerza tras la detección en Estados Unidos de la operación *MoonLight Maze*, considerada la primera gran ciberoperación rusa contra sistemas críticos norteamericanos en el que se accedió a datos militares de gran valor (Gragido, & Pirc, 2011), en unos tiempos donde la seguridad informática casi no existía. Hasta 2017 no se supo del todo qué y cómo se consiguió, al encontrar unos investigadores un año antes el servidor desde donde se había lanzado el ataque (Guerrero-Saade, & et al., 2017). Simultáneamente, a partir de esos hechos, tanto China como Rusia y EE.UU. ratificaron en sus doctrinas de seguridad el estatus de la información digital como un medio ofensivo equiparable a los misiles en capacidad destructiva y táctica, lo que aceleró la carrera por el desarrollo de la ciberguerra y la defensa informativa.

1.1.6. El ataque a las torres gemelas de Nueva York

Tras los atentados del 11 de septiembre de 2001 contra las Torres Gemelas de Nueva York, el sector de la inteligencia y la seguridad en Estados Unidos y otros países occidentales experimentó una transformación profunda, dando lugar a

una "economía de la vigilancia" en la que la extracción, análisis y correlación masiva de datos personales se convirtieron en la columna vertebral de la estrategia antiterrorista. Se desarrollan a partir de ese momento herramientas basadas, por ejemplo, en la teoría de redes, empleada para analizar redes sociales y detectar matemáticamente a los grupos y las cuentas más influyentes, o que conecten grupos entre sí (Moncrieff, Kilibarda, & Gaggioli, 2024).

Empresas tecnológicas como *Palantir Technologies* (Palantir es la bola donde Saruman, mago de El Señor de los Anillos podía ver y controlar a distancia, y cuyo cofundador fue Peter Thiel, cofundador de Paypal junto a Elon Musk, propietario de las empresas Tesla o X, entre otras) emergieron del clima post-11S, desarrollando plataformas capaces de integrar y analizar grandes volúmenes de datos (*big data*) con el objetivo de detectar redes de amenazas, cruzar información de fuentes diversas y modelar escenarios de riesgo antes inimaginables. Palantir, fundada en 2003 con el apoyo inversor de la CIA a través de In-Q-Tel, fue utilizada por agencias gubernamentales estadounidenses para inteligencia criminal, militar y de seguridad nacional, y su popularidad se trasladó a mercados de defensa y sector privado internacional (QuantumSEC, 2024). Plataformas como *Analyst's Notebook* de i2 Group permitieron visualizar redes de conexiones complejas en tramas criminales, y la aplicación de este tipo de análisis extendió masivamente el uso del Análisis de Redes Sociales (SNA) para terrorismo, espionaje y crimen organizado.

El periodo también fue testigo de un auge de empresas de ciberseguridad, vigilancia electrónica y biometraje, estimuladas por la extensión de la *Patriot Act* y la legalización de la recopilación masiva de metadatos telefónicos y digitales. Programas como PRISM y acuerdos secretos con grandes empresas tecnológicas transformaron la privacidad digital y permitieron extraer información estratégica desde las redes sociales. El acceso a datos financieros internacionales (programa TFTP/SWIFT), el uso de información bancaria, la explotación de cámaras de vigilancia y reconocimiento facial, y la colaboración de gigantes de telecomunicaciones completaron un círculo que permitió a gobiernos y empresas asociadas una capacidad inédita para mapear, predecir y perseguir amenazas percibidas (Gellman, & Poitras, 2013).

La consolidación de este "nuevo complejo industrial de la vigilancia" generó asimismo un intenso debate social y político sobre los límites entre seguridad y derechos fundamentales. Revelaciones como las de Edward Snowden sobre los programas de análisis de redes sociales de Estados Unidos en 2013 pusieron de manifiesto el alcance de la vigilancia masiva y obligaron a una reeva-

luación de las prácticas de inteligencia, provocando reformas parciales en la legislación, mayor supervisión parlamentaria y nuevos movimientos en defensa de la privacidad (Snowden, 2019).

1.1.7. Estonia en 2007

La crisis del soldado de bronce en Estonia en 2007 (Soldatov, & Borogan, 2015; Jankowicz, 2020) representa uno de los primeros ejemplos en las guerras híbridas modernas y la historia de la ciberseguridad internacional. Tras la decisión del gobierno estonio de trasladar una estatua conmemorativa de los caídos soviéticos, se produjeron disturbios violentos protagonizados principalmente por la minoría rusa y amplificados por la narrativa de medios rusos, llevando a saqueos, decenas de heridos, más de mil detenidos y la muerte de un manifestante.

El acto simbolizó el choque de memorias históricas, ya que para muchos estonios, el monumento era un recuerdo de la ocupación soviética y suprimirlo suponía reafirmar la identidad nacional. En cambio, para Rusia y muchos ciudadanos rusoparlantes, era un símbolo de la victoria sobre el nazismo y su traslado fue percibido como una afrenta. La reacción fue inmediata, tanto política como mediática, y culminó en un ciberataque de denegación de servicio (DDoS) masivo, dirigido contra sitios gubernamentales, bancos, medios y servicios críticos de Estonia. El país, conocido como "e-Stonia" por su desarrollo digital, quedó parcial o totalmente aislado digitalmente durante semanas, estableciendo así el primer precedente de ciberataque coordinado atribuido a una motivación estatal y geoestratégica.

Este episodio demostró la vulnerabilidad de las infraestructuras digitales ante ofensivas híbridas que combinan protesta social, manipulación mediática y acción cibernética disruptiva. Aunque nunca se pudo demostrar formalmente la implicación directa del Kremlin (en parte por la dificultad estructural de atribuir ataques cibernéticos y la presencia de actores "patriotas" no oficiales), la crisis marcó el inicio de una era en la que los Estados comprendieron que los ciberataques podían utilizarse como armas estratégicas, anticipando patrones que se replicarían posteriormente en Georgia (2008) y Lituania (2008).

Como consecuencia, la OTAN decidió crear y ubicar su Centro de Excelencia de Ciberdefensa Cooperativa en Tallin, capital de Estonia, simbolizando el reconocimiento internacional de la importancia geopolítica de la ciberseguridad y de la defensa integral frente a operaciones de guerra híbrida.

1.1.8. La Operación Plomo Fundido de Israel

La operación "Plomo Fundido" (diciembre 2008-enero 2009) marcó otro paso en la guerra informativa y digital contemporánea, ejemplificando cómo el conflicto militar se entrelazó con la gestión estratégica de la narrativa en espacios digitales. Ante la cobertura internacional predominantemente crítica tras el inicio de los ataques sobre Gaza, Israel detectó que perdía la batalla por la legitimidad del relato en los medios tradicionales, y por ello apostó de manera decidida por la diplomacia pública digital y la propaganda proactiva.

Las Fuerzas de Defensa de Israel pusieron en marcha una innovación a través del canal oficial de YouTube, donde difundieron imágenes selectas de sus operaciones, intentando de manera simultánea mostrar precisión militar y la minimización de daños colaterales civiles (Rodríguez-Fernández, 2024). Paralelamente, desplegaron una estrategia intensiva en Twitter y otras redes, utilizando varios idiomas, *hashtags* y recursos multimedia para justificar la intervención, contestar en tiempo real los argumentos del grupo palestino Hamas y viralizar su propio marco narrativo (Caldwell, & et al., 2009). La campaña de *Hasbará* ("explicación" en hebreo) movilizó tanto portavoces oficiales como redes de ciudadanos y estudiantes, quienes respondían activamente en foros, redes, debates y hasta editaban entradas de Wikipedia, logrando una influencia coordinada y distribuida sobre la imagen internacional de Israel (Aouragh, 2016). Desde entonces, la *Hasbará* se afianzó como modelo de diplomacia pública israelí, evolucionando hacia operaciones híbridas de influencia y desinformación digital asistida por inteligencia artificial y coordinación entre actores estatales, civiles y tecnológicos (Yair, & Perlov, 2024).

1.1.9. Las primaveras árabes, el 15M y Occupy Wall Street

Entre 2010 y 2012, la irrupción de las Primaveras Árabes, el 15M español y *Occupy Wall Street* en Nueva York marcaron el inicio de una era de protesta digital global, en la que las redes sociales y la conectividad móvil transformaron radicalmente la capacidad de las multitudes para organizarse, transmitir información y desafiar regímenes autoritarios o el orden económico establecido. Las protestas de Túnez, Egipto e Irán evidenciaron que la viralización de vídeos testimoniales y la coordinación a través de plataformas como Twitter y Facebook permitían eludir la censura, movilizar a la sociedad y presionar a la comunidad internacional. Al mismo tiempo, fenómenos como el 15M en España (Arévalo Salinas, 2014) o el movimiento *Occupy Wall Street* mostraron cómo la "organización abierta", sin líderes claros y con exigencias difusas, podía potenciar tanto

la participación colectiva y la creatividad como dificultar la negociación con los poderes establecidos (Beran, 2019).

El caso de Egipto, donde el régimen del entonces presidente Hosni Mubarak llegó a cortar el acceso total a Internet y redes móviles, ilustra la importancia vital que había adquirido la conectividad digital para la movilización ciudadana y la visibilidad internacional del conflicto. En contextos como Túnez, Marruecos e Irán, la viralización de imágenes y testimonios, la proliferación de plataformas de "periodismo ciudadano" y el uso de Twitter para coordinar estrategias de protesta globalizaron la denuncia y aceleraron la caída de regímenes, modificando la percepción sobre la legitimidad de los gobiernos y el apoyo que podían recibir desde el exterior.

El impacto de este ciclo no sólo inspiró tácticas imitadas entre movimientos, sino que inauguró una era de tecnopolítica "de código abierto", en la que los activistas podían participar y redefinir los objetivos de manera descentralizada. El resultado fue una nueva forma de organización donde, como en *Occupy Wall Street*, la ausencia de demandas claras ofrecía tanto fortaleza como vulnerabilidad comunicativa, haciendo difícil su éxito y diluyendo su impacto negociador (Colon, 2021). El catedrático de sociología y que fue Ministro de Universidades en España en el periodo 2020-21, Manuel Castells, llegó a exponer en su obra "Comunicación y Poder" (2009) cómo las redes habían venido a democratizar a la sociedad, ya que a través de ella se podía transformarla y hacerla participativa.

Este ciclo de movilización digital influyó profundamente a nivel internacional, generando en ciertas élites de diversos países la convicción de que Occidente estaba utilizando la "revolución tecnopolítica" como arma estratégica. Las elecciones legislativas rusas de 2011 se convirtieron en el punto de inflexión, pues tras ser denunciadas como fraudulentas por la OSCE y por la Secretaria de Estado de Estados Unidos Hillary Clinton, el régimen percibió las protestas masivas, especialmente del opositor Alekséi Navalni (que sería envenenado en 2020 y moriría en 2024 en una cárcel rusa) y la aparición de versiones locales de "*Occupy*" como una amenaza existencial promovida desde el exterior, lo que llevó al Kremlin a reinterpretar la protesta social como parte integral de una supuesta guerra híbrida occidental (Soldatov, & Borogan, 2015). A partir de 2012, el Kremlin blindaría el ecosistema digital nacional e iniciaría el proceso de desconexión del resto de internet del mundo, anunciándose a finales de 2025 que en 2026 se terminaría de completar (Cuesta, 2025).

1.1.10. El libro de Dugin

La figura de Alexander Dugin, teórico y filósofo ruso, es fundamental para entender la articulación ideológica, estratégica y propagandística de la política rusa contemporánea, especialmente en su dimensión digital y mediática. Autor del libro "Los fundamentos de la geopolítica" en 1997 y principal promotor del neoeurasianismo sostuvo que la guerra informativa, la manipulación mediática y la propaganda digital son herramientas cruciales en la lucha geopolítica entre Rusia y Occidente (Tsygankov, & Tsygankov, 2021). Sus postulados han permeado la lógica estratégica rusa, donde la fragmentación de la opinión pública, la desestabilización discursiva y la construcción de comunidades polarizadas se han convertido en tácticas recurrentes adoptadas por los operadores rusos en la guerra digital, tanto en la invasión de Crimea de 2014 como en campañas de desinformación en Europa y Estados Unidos.

El libro articula la visión de una Rusia poderosa y euroasiática, enfrentada al liberalismo occidental, y define las operaciones informativas y psicológicas como parte central de la expansión geopolítica. Dugin plantea que la guerra no solo debe librarse en el terreno militar, sino también en la narrativa, la cultura, la identidad y la percepción, situando la manipulación de la opinión y la producción de propaganda como armas de primer orden en el arsenal ruso. De esta manera Dugin propone como principales ideas (Colom Piella, 2020):

- Creación de un bloque euroasiático liderado por Rusia para contrarrestar el poder de Estados Unidos y la OTAN. “Finlandización” de Europa como modelo de sumisión sin ocupación directa, separación del Reino Unido del resto de Europa, anexión de Ucrania, alianza con Irán frente Occidente, equilibrio y a la vez negación de la expansión de China como rival estratégico (Galeotti, 2014).
- Plantea la importancia de una guerra informativa y psicológica para desestabilizar sociedades adversarias y fortalecer la cohesión interna mediante la manipulación mediática, ideológica y cultural.
- Defiende una visión conservadora y tradicionalista, contraria al liberalismo, promoviendo la restauración de valores y estructuras "orgánicas" frente a la modernidad globalizadora.
- Reafirma el uso del nacionalismo y la identidad colectiva como mecanismo central de movilización social y política.

Su influencia política e ideas, aunque no es directa en el aparato del Estado ha sido decisiva en la filosofía rusa y también ha trascendido a su país, alimentando movimientos identitarios y nacionalistas en Europa a través de redes de la "Nueva Derecha" (Cubero Trujillo, 2020), consolidando la interrelación entre discurso radical, estrategia digital y manipulación algorítmica.

1.2. LA CULTURA DIGITAL Y SU INFLUENCIA EN LA SOCIEDAD

1.2.1. La cultura digital y los videojuegos

En la década de 1990, la cultura digital y la sociedad occidental vivieron una doble paradoja, donde por un lado estaba la euforia del progreso tecnológico y la proliferación de mundos fantásticos, facilitados por el crecimiento de la televisión, los videojuegos y, sobre todo, el acceso a internet. Por otro, el surgimiento de una profunda sensación de desencanto y desencuentro con las promesas de movilidad social, revolución cultural o emancipación que la generación anterior, la de los años sesenta, había cultivado. En el contexto estadounidense, el capitalismo parecía haber alcanzado una posición inamovible, potenciado por el consumismo y la mercantilización de la rebeldía juvenil, donde la música *grunge* y el *hip-hop*, por ejemplo, pasaron de ser catalizadores de crítica y descontento social a convertirse en vehículos de venta de estilos de vida y productos. La contracultura, lejos de continuar como motor de cambio, se encontraba atrapada entre el vértigo de la multiplicación de pantallas, la nostalgia del propio pasado y el agotamiento de un relato de esperanza en el futuro. La evidencia más visible fue quizás la proliferación de universos de escape, tanto en el entretenimiento como en la incipiente vida *online*, una tendencia reforzada por eventos simbólicos como la tragedia de Woodstock '99, que concluyó en caos, violencia y explotación comercial, exponiendo el naufragio de los viejos relatos de paz y amor.

Esta cultura de la insatisfacción y el escepticismo propició el nacimiento de nuevas formas de asociación y diálogo en los primeros foros y comunidades digitales. Los usuarios, hijos de una generación acostumbrada a consumir pero también a desconfiar de los discursos dominantes, encontraron en la red un laboratorio para la ironía, el absurdo y la transgresión, así como una válvula de escape ante el aislamiento social o el descrédito de las instituciones. Frente a la percepción de un mundo estático y sin futuro, internet era, a la vez, un refugio y un terreno para la experimentación con identidades, discursos y códigos cada vez más desligados de la realidad convencional. Así surgieron los primeros

grandes foros de nicho (BBS, *Usenet*, *Something Awful*), y luego, en Japón, los foros como 2channel, que sentaron las bases para el anonimato, la viralidad y el humor corrosivo. La figura del *otaku* japonés (aislado, absorto en universos paralelos y subculturas), pronto se replicó en el usuario estadounidense medio, desencantado y cínico ante la promesa de cambio social, pero muy creativo en la elaboración de formas de símbolos de resistencia.

Entre los años 2000 y 2010, este paisaje cultural y tecnológico sufrió una transformación acelerada que tendría consecuencias profundas para la política, la sociabilidad y la cultura colectiva occidental. La evolución de los foros en Estados Unidos (con la web *Something Awful* como ejemplo de la sátira donde nada tiene sentido, fenómenos paranormales y el humor que rozaba lo grotesco), se entrelazó con el surgimiento de nuevas plataformas mucho más permisivas, anónimas y caóticas, donde la más influyente sería la red social 4chan. Esta web, nacida del espíritu de los foros japoneses, difundió muchas de las dinámicas de anonimato absoluto, ausencia de reglas más allá de la legalidad, rivalidad humorística permanente y una inclinación a transgredir cualquier frontera ética o de gusto, ya fuese por pura diversión, por catarsis, o por construir una identidad colectiva basada en la ironía y la provocación.

Lo que caracteriza este periodo no es sólo la génesis y consolidación de lo que hoy conocemos como "subculturas" digitales (*gamers, trolls, incels*, etc.) sino una nueva lógica en la que los conflictos y las normas sociales podían ser creados, destruidos y reinventados en tiempo real, a golpe de broma, meme u ofensiva colectiva (los llamados "*raids*"). La "democratización" de la creatividad y la comunicación, sin embargo, coexistió pronto con la aparición de dinámicas mucho más oscuras, así la misoginia y el racismo, por ejemplo, encontraron una cultura del troleo e ironía que legitimaba el extremismo bajo el pretexto de la transgresión. Las narrativas conspirativas, las comunidades de autoayuda masculina y la cultura del foro como refugio de los descontentos, comenzaron a entrelazarse en un ciclo de autorreferencias y endogamia que rápidamente se volvió tóxico, sobre todo a raíz de acontecimientos como la masacre en la escuela secundaria de Columbine en 1999 o los ataques del 11 de septiembre en Nueva York de 2001. Estos hechos incentivaron la retirada colectiva hacia lo digital, la desconfianza hacia la realidad y la búsqueda de nuevos relatos alternativos.

Durante estos años, el escepticismo respecto al futuro se convirtió en un rasgo estructural de la juventud global. El fenómeno fue acompañado por una ansia inusitada de simulacros, estéticas retro, bucles de nostalgia y una creciente propensión a la depresión, la ansiedad y la precariedad, en consonancia

con la crisis económica, la explosión de las redes sociales y la evidente incapacidad institucional para reconectar con la sociedad. Así, los foros y las primeras redes sociales funcionaron a la vez como refugio del desencanto y de la creatividad ilimitada, pero también como laboratorio de experimentación para formas embrionarias de cultura política y guerra simbólica que, al acabar la década, estaban listas para saltar fuera de la pantalla y transformar la dinámica pública internacional. El resultado sería una configuración cultural y política inédita donde una generación huérfana de relato colectivo, pero dotada de un arsenal de memes, narrativas y tecnología sin parangón en la historia reciente, estaba lista para lanzarse (quizás sin tener plena conciencia de ello) a la gran batalla por el control y la manipulación del significado en la era digital.

1.2.2. 4chan

La historia de 4chan es, en muchos sentidos, el relato del auge y la transformación de la contracultura digital en el siglo XXI. Fundado en octubre de 2003 por Christopher Poole, alias "*moot*", un adolescente estadounidense fascinado por los foros anónimos de imágenes japoneses (como 2channel), su objetivo inicial era sencillo: crear un espacio donde él y sus amigos pudieran compartir imágenes y discutir sobre anime y cultura "*otaku*". Sin embargo, el experimento desbordó cualquier expectativa, ya que gracias a una interfaz extremadamente sencilla, anonimato absoluto (cualquiera podía postear sin registro y todos eran, en esencia, "*Anonymous*") y la libertad casi total para publicar cualquier cosa. 4chan se transformó rápidamente, en cuestión de meses, en un hervidero de participación y viralidad transgresora (Beran, 2019).

Los usuarios podían soltar cualquier tipo de ocurrencia, provocar, experimentar, crear memes y subvertir discursos sociales sin mayores consecuencias (o eso creían). En poco tiempo, el sitio se convirtió en el centro neurálgico de la creación y reciclaje de memes, cultura pop distorsionada, campañas coordinadas de broma, y una competencia feroz por ser el más ingenioso, perturbador o inventivo. 4chan no sólo fue la cuna de memes esenciales como LOLcats, Rickroll, la rana Pepe (*Pepe the Frog*, adoptada posteriormente como símbolo de la extrema derecha americana) y muchos más, sino que también sentó los cimientos de una nueva forma de entender (y manipular) la cultura digital en tiempo real (Donovan, Dreyfuss, & Friedberg, 2022).

A partir de mediados de la década de 2000, la comunidad de 4chan, y en especial su tablero más caótico ("/b/"), fue refinando una cultura de la ironía y la transgresión donde no existían tabúes y cualquier límite era motivo de parodia. A

la vez, empezaron a brotar dinámicas más oscuras, apareciendo comunidades enteras de trolls (personas que se hacen pasar por otras en las redes) dedicados a acosar, sabotear y ridiculizar a personas, plataformas o movimientos sociales, con una ética interna que oscilaba entre la nada y el cinismo radical. Sin embargo, de este caldo de cultivo también surgieron movimientos novedosos y contradictorios como por ejemplo, *Anonymous*, una colectividad nacida en 4chan que evolucionó hacia el hackeo informático global, organizando desde ataques de broma hasta protestas reales en las calles contra la Iglesia de la Cienciología en 2008 (donde el uso de la máscara de Guy Fawkes, de V de Vendetta, se volvió icónico) o en favor de la libertad de información. Por otro lado, la progresiva deriva de ciertos tablones hacia discursos de odio, misoginia y racismo (especialmente en el tablero "/pol/" o *Politically Incorrect*), fue el preludio de fenómenos de radicalización digital y guerra cultural, donde estos episodios revelaron la capacidad de transformación del foro, pasando de espacio para la mofa intrascendente, a catalizador de movilización social local y global (Elley, 2021).

Sin embargo, conforme la comunidad se radicalizaba y surgían tablones con tendencias misóginas, nacionalistas y conspirativas, 4chan dejó de ser solo cuna de bromas para derivar en un semillero de campañas de intoxicación política y, en casos extremos, de violencia física real. El bulo *Pizzagate* en Washington D.C. derivó en 2016 en que un hombre armado asaltara una pizzería, convencido por las teorías conspirativas de la existencia redes de abuso infantil en los sótanos de una pizzería (que ni siquiera tenía sótano). Más tarde, ataques aún más graves como el atentado a 51 personas en una mezquita en Christchurch (2019) en Nueva Zelanda (Every-Palmer, & et al., 2020), encontraron en 8chan (heredero y lugar aún más extremo que 4chan) un espacio donde los autores publicaban manifiestos racistas y retransmitían sus crímenes en tiempo real, buscando la repercusión viral. El atentado de Christchurch ha sido reproducido posteriormente en videojuegos, como en ciertos servidores de Minecraft (Gagandeep, 2025), o quiso ser imitado en un atentado en España, detenido antes de cometerse por la Policía en noviembre de 2025 (Maestre, 2025).

Lo más impactante vino después de 2010, con la politización explícita del sitio, ya que usuarios de 4chan desempeñaron un papel clave en la gestación de campañas como *Gamergate* (que se explica más adelante) o el auge del movimiento *alt-right* (extrema derecha americana), utilizando y perfeccionando la guerra de memes y la intoxicación informativa para influir en debates públicos y procesos electorales, incluyendo la campaña de Donald Trump de 2016. La herencia y las consecuencias de 4chan son indudables, ya que es imposible enten-

der la cultura de hoy de internet, la dinámica de los memes, las guerras simbólicas y las actuales crisis de desinformación sin mirar de frente al espejo oscuro (y, a veces, cómico) que este foro ha ofrecido al mundo contemporáneo.

1.2.3. Steve Bannon

Steve Bannon fue una pieza central en la convergencia entre los usuarios de 4chan y la nueva derecha radical estadounidense, especialmente hasta el ciclo electoral que llevó a Donald Trump a la presidencia en 2016. Bannon, presidente ejecutivo de la web de noticias *Breitbart News*, veía en los jóvenes usuarios de foros anónimos (4chan y luego 8chan) un potencial ejército de activistas digitales, desencantados, con destrezas en la manipulación narrativa y libres de la corrección política que predominaba en plataformas más reguladas. Esta visión provino de los ataques por parte de esos jóvenes hacia sus empresas en Asia que vendían adelantos y utensilios en videojuegos a cambio de dinero.

Su medio de comunicación, *Breitbart News*, es un medio digital estadounidense de noticias y opinión de extrema derecha, fundado en 2007 por el comentarista conservador Andrew Breitbart como proyecto "pro-libertad" y pro-Israel, concebido como una suerte de versión derechista del Huffington Post. Con sede en Estados Unidos, se convirtió en plataforma clave de la *alt-right* y en un nodo mediático central del Trumpismo, acumulando críticas por difundir contenidos sensacionalistas, teorías conspirativas y piezas señaladas como misóginas, xenófobas y racistas por distintos analistas y medios (Wendling, 2016; Lever, 2016). En noviembre de 2025 el vicepresidente de EEUU, J.D. Vance, propondría a Jeff Bezos, propietario del diario *The Washington Post*, que pusiera al frente de dicho periódico al corresponsal de Breitbart en la Casa Blanca (Gardner, 2025).

A lo largo de la década de 2010, Bannon y sus aliados (como Milo Yiannopoulos, bloguero de extrema derecha famoso por sus discursos contra el islam, feminismo y la corrección política) identificaron la energía rompedora, la creatividad de los memes y la ironía de estos usuarios no como un simple desahogo, sino como una herramienta poderosa para influir en la cultura y la política real. En ese entorno no solo estudió los modos y lenguajes de la subcultura "chan", sino que activamente los incorporó a la estrategia mediática de la *alt-right*. Su contacto con los *gamers* del escándalo *Gamergate* fue crucial, pues fue donde Bannon comprendió que podía canalizar el resentimiento y la creatividad de estos colectivos hacia objetivos políticos concretos, particularmente por medio de narrativas *antiestablishment*, teorías de la conspiración y una agresiva comunicación orientada permanente al conflicto. Bajo su

influencia, Breitbart y otros medios aliados amplificaron y normalizaron símbolos, memes y eslóganes nacidos en 4chan, hasta llevarlos al corazón de las campañas electorales y la conversación pública.

La relación, sin embargo, fue compleja. Usuarios de los foros a veces veían a figuras como Bannon y los ideólogos más visibles del *alt-right* como "cómplices útiles" o incluso como oportunistas, arriesgándose a traicionar el anonimato, la ironía y la irreverencia original del mundillo chan. Durante la campaña de Trump y muy especialmente tras su victoria en 2016, emergieron conflictos entre parte del mundo chan y los supremacistas clásicos y otros sectores radicalizados. De esta manera, tras el "*Unite the Right*" de Charlottesville en 2017 (en un evento convocado por grupos de extrema derecha y supremacistas blancos para oponerse a la retirada de la estatua del general confederado Robert E. Lee, finalizó con violentos enfrentamientos y un atropello intencionado que causó la muerte de una persona contramanifestante así como decenas de heridos, hechos a los que Trump definió como violencia debida "por muchos lados") (BBC News, 2017), algunos usuarios comenzarían a abandonar 4chan hacia nuevas plataformas aún menos reguladas.

1.2.4. El Gamergate

El *Gamergate* fue una campaña de acoso masivo y polarización que estalló en 2014 y marcó un antes y un después en la cultura digital y en el debate público sobre videojuegos, género y poder en internet. Todo comenzó cuando la desarrolladora de videojuegos Zoe Quinn fue acusada públicamente por su expareja de haber mantenido relaciones con un periodista de videojuegos (acusación que nunca se probó) a cambio de cobertura favorable para su juego "*Depression Quest*" en una revista. Aunque la acusación era infundada, encendió la mecha en los foros de 4chan y otros espacios, donde se dio inicio a una auténtica ofensiva coordinada de ciberacoso, amenazas y desinformación, focalizada originalmente contra Quinn. Muy pronto se expandió contra otras mujeres desarrolladoras, con críticas a feministas, periodistas y figuras que abogaban por la diversidad e inclusión en la industria.

Lo que hizo único al *Gamergate* no fue solo la escala del hostigamiento (*doxing*, amenazas de violación, amenazas de muerte, extorsión y persecución personal) sino la manera en que se organizó y propagó, donde cuentas anónimas de Twitter y campañas coordinadas en redes sociales y plataformas como Reddit, 4chan y 8chan. Aunque sus promotores alegaban que el movimiento era una protesta contra la supuesta corrupción en el periodismo de videojuegos, en la

práctica fue un fenómeno de reacción violenta contra la presencia y la influencia de mujeres, personas LGTBQ+ y voces críticas dentro del sector que se aproximó a lo que sería el futuro movimiento MAGA (*Make America Great Again*, lema del presidente Trump) (Donovan, Dreyfuss, & Friedberg, 2022).

Gamergate se convirtió, de hecho, en una verdadera guerra cultural, pues para sus defensores, era una "lucha por la pureza del *gaming*" y contra el progresismo percibido como intrusión externa. Por contra, para sus víctimas y críticos representó el despertar brutal de la misoginia estructural y la toxicidad latentes en la comunidad de los videojuegos desde hacía años. El movimiento instauró estrategias de manipulación mediática y radicalización que después serían empleadas en otras batallas culturales digitales, sirviendo además de antesala para la "guerra de memes" y el tipo de campañas de desinformación política que se popularizarían en los años siguientes. Incluso fue visto como una "oportunidad soñada" por parte de Steve Bannon para reclutar a jugadores en la batalla contra las élites (da Empoli, 2019).

A nivel cultural y social, *Gamergate* fue el punto de inflexión que traspasó las fronteras de la industria de los videojuegos, marcando el salto de las guerras culturales digitales al debate público general, e influyendo en el desarrollo posterior de la nueva derecha digital y la polarización de las redes sociales.

1.2.5. 8chan

8chan, también conocido como *Infinitechan* o *Infinitychan*, fue creado en octubre de 2013 por Fredrick Brennan, un joven programador estadounidense conocido en la red como "*Hotwheels*". La motivación de Brennan para crear 8chan surgió de su descontento con las crecientes restricciones y censuras que observaba en 4chan, el foro que hasta ese momento era el epicentro de la cultura anónima y de memes en internet. Mientras experimentaba un viaje de hongos psicodélicos, Brennan concibió la idea de un foro radicalmente orientado a la "libertad de expresión", donde los propios usuarios que fueran expulsados de 4chan migraran a 8chan buscando un espacio sin restricciones, convirtiéndolo en uno de los enclaves favoritos de las nuevas guerras culturales *online* (Beran, 2019; Ollero, 2019).

En 2014, conforme el sitio crecía y los desafíos legales y técnicos se multiplicaban, Brennan se asoció con Jim Watkins, un veterano del ejército estadounidense residente en Filipinas, quien asumió el control logístico y, más tarde, la propiedad de 8chan. Brennan siguió vinculado a la gestión hasta 2016, momento en el que se apartó completamente del proyecto y, con el tiempo, se ha convertido en un fuerte crítico del mismo y de su rol en la propagación de discursos

de odio y teorías de la conspiración. 8chan fue cerrado tras la matanza de El Paso en 2019 (en el que un supremacista blanco mató a 23 personas y 23 heridas en un supermercado Walmart, tras anunciarlo en el foro 27 minutos antes) (Barroquere, 2023), y ha estado implicado en la difusión de contenidos extremadamente violentos, como manifiestos de tiroteos masivos y propaganda supremacista, lo que ha provocado periodos de cierre, fuertes debates sobre la responsabilidad digital y la migración de sus usuarios a foros aún menos regulados. Los usuarios de 8chan emigraron posteriormente a 8kun, cuyos servidores se ubicaron en Vladivostok, Rusia (Colon, 2025).

1.2.6. QAnon

El movimiento QAnon surgió en octubre de 2017, cuando un usuario anónimo ("Q", autoproclamado como *Q Clearance Patriot*) publicó en el tablón /pol/ de 4chan mensajes crípticos que afirmaban, entre otras cosas, que Hillary Clinton sería arrestada en breve y que se avecinaba una "tormenta" que desenmascararía una red global de élites pedófilas y corruptas, con Donald Trump como héroe destinado a salvar a Estados Unidos. Q se presentaba como un supuesto alto funcionario con acceso a secretos de Estado, y desde el inicio cultivó un clima de misterio, desafíos y "miguitas" (*drops*) diseñadas para que sus seguidores "investigaran" y se sintieran parte de una auténtica operación de inteligencia paralela. Aunque las primeras predicciones nunca se cumplieron, el formato participativo, especialmente en foros como 8chan y luego 8kun, generó una comunidad expansiva de "soldados digitales" dedicados a interpretar, traducir y viralizar sus mensajes a múltiples plataformas, como YouTube, Twitter, Facebook, Instagram y Telegram (Donovan, Dreyfuss, & Friedberg 2022).

QAnon formó parte del ecosistema conspirativo norteamericano y global, pero supo articular mitos de larga duración (como el pánico satánico, los miedos al "Nuevo Orden Mundial" y el "*deep state*" o estado profundo) en una narrativa nueva adaptada al siglo XXI digital. El movimiento fusionó elementos del bulo Pizzagate, escándalos por pedofilia y prostitución como los de Harvey Weinstein y Jeffrey Epstein, y la polarización creciente desde la llegada de Trump al poder, alimentándose de la ansiedad, la desconfianza y la sobrecarga informativa de la era de las redes sociales.

Una de sus claves no fue solamente la narrativa delirante, sino la viralidad y la apropiación de los memes. QAnon pasó de 4chan a 8chan y 8kun por disputas internas y necesidad de menos moderación, y sus adeptos colonizaron comunidades de Facebook, Instagram y hasta círculos de bienestar y maternidad,

modificando su estética y mensajes ("Pastel QAnon") para captar nuevas audiencias durante la pandemia, mezclándose con el movimiento antivacunas y campañas como "*Save the children*" para supuestamente salvar a los niños del uso de ponerse mascarillas en época de pandemia por covid-19. Uno de sus principales relatos pasa por utilizar el lenguaje de la pastilla roja (*red pill*) que hace despertar del mundo Matrix (haciendo referencia a la película de dicho nombre), un mundo en el que las élites nos harían vivir de forma engañada sin saberlo (Bank, Stack, & Victor, 2018).

QAnon también tuvo un rol clave en la radicalización y movilización de una parte considerable del electorado pro-Trump, apoyando campañas como *Stop the Steal* (parad el robo) en 2020 y participando en el asalto al Capitolio del 6 de enero de 2021. Aunque, tras la derrota de Trump, la actividad de Q desapareció y las grandes plataformas empezaron a restringir cuentas relacionadas, la esencia del movimiento sobrevivió dispersándose en redes alternativas y foros secundarios, mostrando su capacidad de adaptación y persistencia.

1.2.7. La evolución del método 2010-14

Estados Unidos desarrolló la operación "*Earnest Voice*" como parte de su estrategia global para contrarrestar la propaganda extremista y terrorista, especialmente la relacionada con Al Qaeda y sus filiales (Colon, 2025). Esta operación fue impulsada por el *United States Central Command* (CENTCOM) en Oriente Medio y consistió en la utilización de *software* especializado de "*persona management*" (gestión de personas), gestionado por la empresa de ciberseguridad Ntrepid. Estos sistemas permitían crear y administrar "*sock puppets*" (marionetas de calcetín), falsos perfiles controlados semiautomáticamente que, en las redes sociales, se dedicaban a contrarrestar la propaganda extremista y enemiga, siempre fuera del territorio estadounidense.

Paralelamente, en el ámbito local durante el mandato de Obama, se detectó la influencia creciente de la propaganda terrorista *online*. Se propuso una "sala de crisis de la información" y, en 2011, se creó el *Center for Strategic Counterterrorism Communications* (CSCC). Este comité coordinaba todas las actividades de comunicación exterior del gobierno estadounidense contra el terrorismo y el extremismo violento. Por ejemplo, en 2012, el CSCC lanzó campañas en línea dirigidas especialmente contra el frente Al-Nusra, una filial de Al Qaeda en Siria, aunque con problemas para igualar la cantidad de contenido que producían los grupos extremistas (Stengel, 2019).

El CSCC organizó equipos multilingües para participar en debates *online* en árabe, urdu y somalí, con el objetivo de disuadir a jóvenes de unirse a organizaciones terroristas. Además, estos equipos gestionaban sitios web de influencia cuya afiliación con el gobierno estadounidense solo era visible en lugares muy específicos (apartado "*about us*"). En 2013 y con la aparición del Estado Islámico, la estrategia evolucionó hacia una "guerra de relatos", con campañas activas en Twitter en lengua árabe, donde se intentaba contestar directamente la propaganda yihadista, aunque la burocracia y naturaleza estatal dificultaban la rapidez y eficacia (Liang, 2015).

Entre 2012 y 2014, Daesh revolucionó las dinámicas propagandísticas del terrorismo internacional al desplegar una estrategia de comunicación digital sin precedentes, apodada "Jihad 2.0". Esta incluyó la producción y difusión masiva de videos de alto impacto, cuya calidad audiovisual y teatralidad superaba el estándar previo del extremismo violento. Usando plataformas como Twitter y YouTube, Daesh viralizó *hashtags* como #AllEyesOnISIS y logró que estos sirvieran tanto para reclutar adeptos extranjeros como para sembrar el terror entre la población local. Un aspecto central fue la difusión de videos de ejecuciones, que no solo buscaban atemorizar a sus enemigos, sino también doblegar psicológicamente ciudades enteras provocando que muchos civiles y soldados huyeran de localidades que aún no habían sido atacadas militarmente, tras retransmitir en tiempo real (y poder ver en sus móviles) las imágenes de la brutalidad de Daesh (Lesaca, 2017).

En 2014, el conflicto entre Israel y Hamas durante la operación "*Protective Edge*" evidenció la consolidación de la "guerra de los relatos" en la esfera digital. Ambas partes desplegaron campañas mediáticas, compitiendo en la creación de *hashtags* viralizados y la publicación de imágenes y testimonios emocionalmente poderosos. Hamas impulsó la difusión global de #GazaUnderAttack, mientras Israel activó su maquinaria de diplomacia digital para presentar su narrativa (Colon, 2025). Un fenómeno fundamental asociado fue un surgimiento del "periodismo ciudadano", donde jóvenes palestinos y civiles de Gaza usaron Twitter, Facebook e Instagram para documentar y compartir en tiempo real fotos, vídeos y mensajes sobre la vida cotidiana bajo los bombardeos (Patrikarakos, 2017). El impacto emocional y la viralización de este contenido influyeron rápidamente en la agenda mediática internacional, modificando percepciones globales y convirtiendo las redes sociales en un campo de batalla tan decisivo como el militar (Zeitzoff, 2016).

En el ámbito social, las técnicas de análisis de redes y de determinación de emociones y sentimientos comenzaron a usarse en campañas electorales. Las

presidenciales de Barack Obama en 2008 y 2012 marcaron el inicio de una nueva era en la comunicación política digital al integrar de forma central por primera vez las redes sociales y la inteligencia de datos en la estrategia electoral. En 2008, el equipo de Obama utilizó plataformas como MySpace, Facebook, YouTube, Flickr o Pinterest, con campañas basadas en la recopilación y segmentación de gran cantidad de datos, pudiendo centrarse en mensajes dirigidos a poblaciones determinadas (Wylie, 2020). De esa experiencia surgió la primera "página de entidad" de Facebook para una figura política, cuando el perfil personal de Obama colapsó por la avalancha de solicitudes y se convirtió en una página de fans, modelo que luego se extendió a otras campañas y organizaciones (Kaiser, 2019). La campaña de 2008 se percibió como sofisticada e ingeniosa en su momento, pero con mensajes hoy considerados simples, e incluso vulgares. En 2012 la campaña dio un salto técnico y metodológico, ya que tras la recogida de ingentes cantidades de datos desde Facebook, fue la primera elección presidencial en usar de forma sistemática modelos predictivos algorítmicos de persuasión para identificar indecisos, segmentando mucho más allá de la demografía clásica según intereses, género, temas clave y probabilidad de apoyo.

Paralelamente, el equipo de Obama tuvo que gestionar un flujo intenso de odio racial y amenazas en redes recibidas en su contra, respondiendo con bloqueos y borrados que convirtieron a los voluntarios en un "ejército" de moderadores. Mientras, en el otro extremo, foros supremacistas como *Stormfront* registraban picos históricos de nuevos usuarios tras la victoria de Obama de 2008 (da Empoli, 2020).

1.2.8. Sofisticación 2015-2017

Entre 2015 y 2017, la desinformación digital alcanzó un nuevo nivel de sofisticación, con actores como Daesh perfeccionando su uso de las redes sociales, bots e información multilingüe para expandir su ideología, instrucción operativa y presencia internacional. Sus revistas *online* "Dabiq" y posteriormente "Rumiyah" lograron combinar marketing viral, vídeos impactantes y una estrategia dinámica apoyada en la resiliencia digital. Así, cada vez que se cerraban cuentas o canales Daesh los reemplazaba por otros previamente preparados, manteniendo intacto su flujo de propaganda incluso frente a los intensos esfuerzos de contravigilancia occidental (López Gómez, Mendieta Díaz, & Micó Faus, 2021).

Occidente, por su parte, respondió con la creación de centros de contranarrativa y guerra digital, como la 77ª Brigada británica y unidades especializadas

de la OTAN, aunque su capacidad de reacción fue limitada por la rigidez burocrática y la dificultad para igualar la adaptabilidad de los yihadistas. En paralelo, *hackers* organizados como *Anonymous* lanzaron campañas para delatar, bloquear o sabotear plataformas y cuentas vinculadas al terrorismo, lo que añadió una dimensión civil descentralizada a la confrontación (Rey García, Rivas Nieto, & Sánchez Alonso, 2017).

1.3. CAMBRIDGE ANALYTICA

La historia de Cambridge Analytica (CA), aunque duró unos pocos años, es un estudio principal y muy ilustrativo de cómo la maquinaria de la desinformación puede llegar a alterar un proceso democrático al mercantilizar la conducta humana, transformando la guerra psicológica en una ciencia basada en datos.

El relato se remonta a la fundación de su empresa matriz británica, *Strategic Communication Laboratories* (SCL Group), en 1990. SCL se especializó históricamente en operaciones psicológicas (*psyops*), la rama de la guerra dedicada a influir en el público "hostil" para lograr objetivos estratégicos. Sus trabajos iban destinados al Ministerio de Defensa británico y la OTAN, para identificar y combatir extremismo y radicalización en línea, especialmente en países islámicos. A dicha empresa llegaría en 2013 Christopher Wylie, canadiense que había pasado por varias campañas en varios países (como la de Obama) aprendiendo distintos procedimientos de segmentación de la población. Como director de investigación, Wylie se encontró rápidamente en el corazón del desarrollo de una metodología muy novedosa. Ayudó a elaborar un sistema automatizado de recolección de datos acoplado a un dispositivo de aprendizaje automático basado en redes neuronales algorítmicas, capaz de predecir estadísticamente los comportamientos electorales. Con la creación de la nueva filial, Cambridge Analytica, Wylie y sus colegas combinaron la psicometría (a partir del modelado de la personalidad de los usuarios de internet según modelo OCEAN, del que se hablará más adelante) con el *big data* (grandes cantidades de datos), lo que les permitió la microsegmentación de audiencias con precisión en base a distintas variables personales (Wylie, 2020).

Para construir esa base de datos masiva y detallada, CA aprovechó las casi inexistentes políticas de gestión de Facebook y la falta de supervisión gubernamental. El salto en la historia se produjo en 2014, cuando el Dr. Aleksandr Kogan, un académico de la Universidad de Cambridge y con relaciones con la Universidad de San Petersburgo, fue reclutado creando una aplicación de cuestionario de preguntas (*thisisyourdigitallife*) en Facebook que, a través de la "API de

los Amigos" (ya que no solo extraía los datos de quien hacia el juego sino también de todos sus contactos), extrajo de forma ilícita los perfiles personales de más de ochenta y siete millones de usuarios de Facebook, un acto que supuso una violación directa de las condiciones de servicio de la red social.

Trinidad y Tobago se convirtió en el escenario experimental de SCL Group para probar sus técnicas avanzadas de perfilación psicológica y manipulación de conducta. En 2014, el Ministerio de Seguridad Nacional de dicho país contrató a SCL para el Proyecto Trinité, presentado como iniciativa para identificar posibles delincuentes mediante el análisis de datos. Sin embargo, el verdadero propósito era político, pues el gobierno buscaba utilizar el sistema para predecir el comportamiento electoral y orientar futuras campañas, en lo que internamente se denominó el "Proyecto Minority Report". SCL accedió a datos censales completos y a flujos de telecomunicaciones, logrando espiar búsquedas y hábitos digitales de prácticamente toda la población. AggregateIQ, filial de CA, desarrolló la infraestructura que cruzó datos de Facebook, registros de navegación y demás datos sacados de proveedores de internet, desvelando así información personal y política sensible sin consentimiento.

Las tácticas operativas incluían, además, estrategias para explotar la división étnica entre indios y afrocaribeños, principales bloques políticos del país, generando tensión y desconexión electoral entre los jóvenes. En 2015 SCL desplegó una campaña centrada en fomentar el desapego y la resistencia no hacia el gobierno, sino hacia el propio proceso democrático, a través de grafitis, carteles y activismo digital que desincentivaba el voto entre la población afrocaribeña. El objetivo era provocar un ambiente de caos y escepticismo, estrategia que Alexander Nix, CEO de CA, describió abiertamente como "muy divertida", aunque generó un gran desorden y polarización social en el país (Kaiser, 2019).

El potencial de llevar estas tácticas al ámbito electoral se hizo realidad cuando el propagandista de la *alt-right*, Steve Bannon, junto con el multimillonario Robert Mercer, fundaron Cambridge Analytica como filial estadounidense de SCL, con una inversión significativa (en torno a 15 millones de dólares). Bannon veía a CA como el "arsenal" necesario para ganar su "guerra cultural" y alcanzar la "dominación de la información".

La visión de CA era inicialmente sencilla, donde Nix aseguraba haber contratado a científicos de datos tan brillantes que podían aislar a individuos y, literalmente, hacerles pensar, votar o actuar de manera diferente a como lo habrían hecho antes a través de la microsegmentación. Esta técnica se basaba en el modelo de personalidad de los Cinco Factores (OCEAN), que clasifica a los individuos según cinco variables (Apertura, Consciencia, Extraversión, Afabilidad

y Neurosis), donde la medición de la neurosis resultaba crucial, ya que se correlacionaba con la probabilidad de que una persona tomara decisiones impulsadas por el miedo. Este análisis psicográfico, combinado con los miles de puntos de datos que CA aseguraba poseer sobre cada adulto estadounidense, le permitía crear una "realidad" para cada votante.

Antes de aplicarlo a las grandes campañas occidentales, CA puso a prueba sus métodos en más países para depurarlos. De esta manera trabajaron en Nigeria en 2015 para, en lugar de promocionar al candidato de Goodluck Jonathan, centrar la estrategia en destruir la candidatura de su oponente (Buhari) mediante la difusión de propaganda de intimidación y rumores. En dicha campaña mantuvieron contactos con la empresa petrolera rusa Lukoil, interesados en el funcionamiento de su método (Cadwalladr, & Graham-Harrison, 2018; Wylie, 2020).

El arsenal definitivo se desplegó en 2016 durante el referéndum del Brexit en el Reino Unido, donde la campaña *Vote Leave* utilizó una empresa subsidiaria de CA, AggregateIQ (AIQ), para eludir límites de gasto legales. AIQ utilizó los modelos psicográficos de CA para enviar más de mil millones de mensajes personalizados a los votantes persuadibles en Facebook. Tras su éxito fue llevado a su objetivo final, en la campaña presidencial de Estados Unidos de 2016, donde CA trabajó en el Proyecto Álamo para Donald Trump, creando miles de campañas publicitarias distintas individuales en Facebook. Más allá de la persuasión para que se votara a favor de Trump, se instaló un dispositivo masivo para disuadir a los votantes demócratas de que fueran al colegio electoral. Periodistas y denunciantes revelaron que CA utilizó tácticas de supresión de voto, focalizándose en afroamericanos en zonas vulnerables, empleando anuncios que explotaban sus miedos para desmotivarlos de ir a las urnas.

Estos modelos correlacionaban los "me gusta" en Facebook con rasgos de personalidad, y aunque su efectividad real en ese momento fue cuestionada, estimándose una precisión del 30% en predicción de personalidad, era superior al 85-90% en características demográficas y políticas (Hindman, 2018), logrando que con campañas dirigidas de desmovilización la participación afroamericana fuera la más baja en 20 años (Stracqualursi, 2020). Se estimó que el 75% de los 400.000 bots detectados en la campaña favorecían a Trump, y aunque representaban solo al 6% de las cuentas totales que participaban en las redes en campaña, suponían el 31% de desinformación (Hernando, 2020).

Sin embargo, el éxito económico y político de la empresa estuvo marcado por la controversia desde el principio. En 2015, el periódico británico *The Guardian* ya alertó sobre el uso ilegal de datos, lo que obligó a CA a certificar falsamente el borrado de información ante Facebook. La situación llegó al límite en

marzo de 2018, cuando las filtraciones del Christopher Wylie, arrepentido de todo lo hecho, revelaron la magnitud del escándalo, pues CA no solo conservaba los datos ilícitos, sino que los podría haber puesto a disposición de terceros, incluyendo agentes extranjeros. El canal de noticias *Channel 4 News* amplificó la crisis al mostrar mediante cámara oculta a Alexander Nix ofreciendo prácticas abiertamente corruptas y estrategias de chantaje (sobornos, cebos sexuales, manipulación) a posibles nuevos clientes. Las consecuencia fueron inmediatas, con investigaciones globales, intervención del FBI, la ICO británica y otras agencias, así como la suspensión de acceso a CA y SCL Group por parte de Facebook.

En mayo de 2018, ambas empresas se disolvieron definitivamente, pero su legado sigue marcando el debate sobre democracia, privacidad y manipulación digital. Su modelo demostró que la combinación de psicología conductista, análisis masivo de datos, inteligencia artificial y algoritmos de aprendizaje automático puede convertirse en un arma de guerra psicológica, capaz de fragmentar el espacio público, explotar debilidades individuales y distorsionar el resultado de elecciones.

1.3.1. Las consecuencias posteriores de Cambridge Analytica

La primera en recibir las consecuencias de todo lo descubierto fue la periodista que lo destapó, Carole Cadwalladr, que recibió denuncias por difamación por parte del donante del Brexit Arron Banks, al haber sugerido que tenía una relación encubierta con el gobierno ruso en relación con financiación electoral. Tras varias instancias y apelaciones, Banks ganó parcialmente por falta de pruebas definitivas de dicha relación, lo que le supuso a Carole tener que pagar en torno a 1,2 millones de Libras en costas (1,36 millones de €) (BBC News, 2023). Igualmente en abril de 2025 los nuevos propietarios de los diarios *The Guardian* y *The Observer*, donde trabajaba, no renovaron su contrato. Durante todo este tiempo recibió campañas de acoso e insultos (Posetti, & et al., 2024).

Tras el cierre de Cambridge Analytica y SCL Group en 2018, los propietarios y antiguos miembros del conglomerado no tardaron en reconvertirse ni en renovar las estructuras de influencia. Paralelamente, el método psicométrico y de explotación masiva de datos ya habían sido difundidos y normalizados en el sector, así otros exintegrantes como Gaby van den Berg fundaron *Emic Consulting*, aplicando sus técnicas para clientes como los ejércitos de Holanda y Canadá. Este último fue públicamente señalado (por medios como *Ottawa Citizen*) por haber realizado operaciones de influencia sobre su propia población, especialmente durante la pandemia por covid-19 (Pugliese, 2020).

Steve Bannon ascendería a asesor presidencial de Donald Trump hasta 2017. Tras abandonar la Casa Blanca, fundaría en Bruselas la organización *The Movement*, con la ambición de crear una coordinadora paneuropea para impulsar la extrema derecha y formar futuros líderes en un monasterio romano en Italia, aunque su influencia se esfumó cuando los apoyos iniciales, como los de Marine Le Pen en Francia, le dieron la espalda.

El trabajo de seguimiento realizado por Wendy Siegelman sobre el destino del personal y la red corporativa de Cambridge Analytica, desde su clausura formal en mayo de 2018 hasta la actualidad, constituye una de las investigaciones más detalladas y esclarecedoras de estas empresas. Siegelman, periodista norteamericana de investigación, ha documentado este fenómeno a través de su portal Newstracs.com, así como en artículos en diarios como *The Guardian*, *Byline Times* y otras plataformas especializadas, desentrañando la compleja red de empresas, directivos y operaciones internacionales que se gestó a partir de la disolución de Cambridge Analytica. Tal como afirmó en 2018 "Cambridge Analytica ha muerto, pero su oscura red está viva y bien".

Así, según sus informaciones, Emerdata Limited emerge como la pieza central del entramado sucesorio de Cambridge Analytica y SCL Group. Empresa registrada incluso antes de la crisis pública de la consultora original, absorbió directivos, accionistas y activos clave, convertiéndose en el principal vehículo de continuidad. Destacan, especialmente, las figuras de Rebekah y Jennifer Mercer, herederas del magnate republicano Robert Mercer (uno de los primeros desarrolladores de la inteligencia artificial, quien financiara tanto *Breitbart News*, SCL, como la campaña de Trump en 2016) (Gold, 2017), y que han ejercido el control directivo y financiero hasta al menos 2024. Alexander Nix, ex director ejecutivo de Cambridge Analytica, también integró el consejo directivo, aunque fue desalojado rápidamente tras el escándalo mediático de la empresa al hacerse públicos sus actos, dejando paso a Alexander Tayler, ex director de datos, quien asumiría el mando operativo. Julian Wheatland, presidente de SCL y encargado de comunicar el cierre de Cambridge Analytica, encabezó la estructura en la etapa crítica de transición.

Las conexiones internacionales resultan particularmente llamativas, donde Siegelman identificó a actores clave como Johnson Chun Shun Ko, empresario vinculado al sector de la seguridad privada y consejero de *Frontier Services Group,* empresa con participación estatal china y cofundada por Erik Prince, el célebre fundador de *Blackwater*, una empresa de mercenarios (La Sexta, 2020). Ko fue sustituido posteriormente por Gary Ka Chun Tiu, consolidando el nexo entre los intereses asiáticos, las redes de inteligencia privada y

las estructuras de datos occidentales. Cheng Peng (Sandy Peng), con importantes vínculos empresariales en China, los Emiratos Árabes Unidos y la red de Erik Prince, completó esta constelación de directivos internacionales que, en palabras de la propia Siegelman, daban lugar a una "estructura corporativa diseñada para la opacidad y confusión" (Siegelman, 2024a).

Sus investigaciones más recientes indican que Emerdata sigue plenamente activa bajo el control de Rebekah Mercer, manteniendo un inusual número de accionistas (32 en 2021) y una actividad principal relacionada con la financiación de defensas legales y la gestión patrimonial de las antiguas sociedades del entorno SCL. Rebekah es una de las comisarias de la Fundación *Heritage* y creadora de *Project 2025* para crear las propuestas que regirían la política de Donald Trump si ganaba las elecciones de 2024 (Siegelman, 2024b), aunque tras ganar las elecciones el presidente ha querido distanciarse públicamente de las mismas (Levien, 2024).

Pero la evolución empresarial del ecosistema va más allá de Emerdata. Siegelman rastreó en detalle la creación de Auspex International, una consultora política fundada por Ahmad Al-Khatib (ex director de Emerdata) apenas un mes después del cierre de Cambridge Analytica. Mark Turnbull, quien figuraba en los vídeos de cámara oculta del canal inglés Channel 4 explicando estrategias de manipulación electoral, asumió la dirección operativa, acompañado por otros exdirectivos de Cambridge Analytica y SCL. Auspex International comenzó a operar en África y Medio Oriente, desplegando los mismos modelos de microsegmentación y análisis de datos, proponiéndose como una consultora "ética".

Este reciclaje de personal y *know-how* se complementó con la aparición de *Dynamo Recoveries Limited*, empresa dedicada a litigar pleitos heredados y mantener separados los activos de las actividades de influencia. Esta continuidad operativa en los procesos fue posible gracias a la segmentación de responsabilidades legales, la preservación de los cuadros directivos y una red financiera y tecnológica transnacional renovada en sus marcas y formas legales. Las conexiones políticas, mantenidas sobre todo a través de los Mercer, Bannon y otros aliados del trumpismo, así como el refuerzo de vínculos internacionales con personas de China, Emiratos Árabes, Rusia y África, habrían permitido incluso una expansión de los métodos originales de influencia y microtargeting político.

La relación con Donald Trump y su entorno es uno de los hilos más consistentes de las investigaciónes de Siegelman, desde la financiación de la consultora por parte de los Mercer hasta la integración de Steve Bannon y Erik Prince en el núcleo duro del trumpismo, el puente entre Cambridge Analytica y el uni-

verso de influencia política y social del presidente norteamericano. Incluso después del escándalo y cierre de la consultora, el entorno empresarial de Trump seguiría nutriéndose de estrategias, tecnología y personal provenientes de la vieja red. Por ejemplo Sandy Peng, relevante por sus conexiones en el ecosistema cripto, fue en 2024 vinculada a *World Liberty Financial*, una iniciativa promovida por Trump sobre criptomonedas y que evidenciaría la vigencia de alianzas con actores del antiguo Cambridge Analytica (Siegelman, 2024c).

Siegelman constata, además, que la transformación de Cambridge Analytica cristalizó en una proliferación de consultoras, *startups* y empresas de la llamada "comunicación estratégica" en toda Europa continental. La reubicación de personal y tecnologías se realizaría casi siempre bajo estructuras legales opacas, filiales internas en el Reino Unido, Irlanda, Suiza u Holanda, y una fuerte presencia multinacional habilitada por la dispersión de sociedades pantalla y la flexibilidad de jurisdicciones. Las nuevas firmas, aunque adaptadas a regulaciones locales, ofrecerían servicios de minería de datos, microsegmentación y ciberseguridad táctica, repitiendo el modelo precedente.

1.3.2. El negocio y expansión de empresas

El negocio en torno al sensacionalismo, la desinformación, el odio y las teorías de conspiración ofrece numerosos ejemplos, siendo uno de los más destacados el de InfoWars, dirigido por Alex Jones. Este medio de EE.UU. potenciaba un modelo lucrativo que se basa en la explotación de emociones y realidades alternativas, impulsado por una ideología claramente ultraderechista que además se convirtió en un importante propagador de noticias provenientes del canal ruso RT (Colon, 2025). InfoWars, sitio donde la información y las conspiraciones se utilizaban como arma y donde su director se autodefinía como infoguerrero (Byung-Chul, 2018), llegó a obtener unos 165 millones de dólares en ingresos entre 2015 a 2018 especialmente en productos de *merchandising* de su propia web, pero también se han vinculado al menos media docena de hechos violentos en la vida real derivados de su contenido (Iriarte, 2025). En 2022 fue condenado a pagar más de 1.400 millones de dólares por difamaciones relacionadas con una masacre de 26 personas (20 de ellas niños), donde Alex Jones aseguró que había sido un montaje (Sánchez-Vallejo, 2025). Esto le llevó a la bancarrota en 2024, obligándole a vender InfoWars, que fue adquirido por la web satírica y crítica *The Onion* con el objetivo de recaudar indemnizaciones a las victimas de Jones (Pichi, 2024).

En las elecciones estadounidenses de 2016, la pequeña ciudad de Veles, en Macedonia del Norte, se convirtió en un polo de producción de noticias falsas en inglés orientadas casi siempre al electorado conservador y favorables a Donald Trump. Era impulsado por jóvenes y pequeños grupos cuya motivación principal era el beneficio económico mediante publicidad programática y tráfico desde Facebook, más que un proyecto de injerencia geopolítica. A partir de cursos y tutoriales de marketing digital, técnicas de SEO, elección de titulares impactantes y segmentación en redes sociales, se llegaron a registrar más de un centenar de dominios que imitaban medios estadounidenses, copiaban y adaptaban contenidos de portales conspirativos de EE.UU. generando millones en ingresos (Hughes, & Cvetkovska, 2021; Allcott, & Gentzkow 2017).

Tras ese ciclo electoral, la infraestructura creada desde Macedonia no desapareció sino que se reconfiguró y reapareció en campañas posteriores, incluidas las presidenciales de 2020 en Estados Unidos, manteniendo el patrón de "*clickbait partidista*" y ajustando las tácticas para ocultar mejor la procedencia real de las páginas mediante identidades falsas y migración hacia plataformas con menor moderación. Paralelamente, Macedonia del Norte se ha consolidado como un terreno de disputa informativa en sí mismo, con portales locales y redes sociales que amplifican narrativas pro-Kremlin sobre la OTAN, la UE o el referéndum de 2018, así como contenidos que encajan en estrategias más amplias de influencia de actores externos en el espacio mediático de los Balcanes (Strategic Analysis, 2023; Dumont, Solis, & Zaleski, 2023).

Este proceso de expansión se convierte ya en algo global, así la Dra. Emma L. Briant señala que el número de empresas dedicadas a propagación psicométrica y desinformación pasó de menos de 20 en 2017 a unas 600 en 2020 (Briant, 2020). El reputado *Computational Propaganda Research Project* de la Universidad de Oxford en su informe "*Industrialized Disinformation*", expuso que en 2020 ya había presencia confirmada de ciber-tropas y empresas especializadas en 81 países (Bradshaw, Bailey, & Howard, 2020). Estos datos exponen de manera muy clara la difusión de empresas de uso de datos e influencia de redes sociales e internet.

El *The Influence Industry Project* (2022), con sede en Berlín (Alemania) lanza un proyecto en el que, en colaboración de periodistas, académicos y organizaciones de más de 20 países, expone a más de 500 empresas y organizaciones que operaron en el mundo y ofrecen servicios de microsegmentación, análisis de datos, consultoría digital y/o manipulación del entorno informativo (desde grandes firmas de marketing político hasta agencias locales que trabajan para partidos, gobiernos y candidatos), mostrando que el negocio de la persuasión política basada en datos es mucho más amplio que los casos famosos como Cambridge Analytica.

En la web del proyecto se ofrece un buscador (https://influenceindustry.org/en/explorer/companies/) en el que permite buscar por países o tipo de servicios, donde se exponen diversas empresas por comportamiento no auténtico, por uso de cuentas falsas para hacer campañas de apoyo masivo y coordinadas, inflando opiniones o informaciones desde países como Egipto, Emiratos Árabes, Ucrania, Indonesia, Israel o Ecuador. En campañas de desinformación identifica empresas de Egipto, Indonesia, Israel o Filipinas. En España el buscador de este proyecto sitúa actuaciones en el propio país únicamente de influencia o inteligencia algorítmica, sin llegar a ser aspectos ilegales, a varias empresas.

1.4. LA EVOLUCIÓN EN RUSIA

En las elecciones legislativas rusas de 2011, el Kremlin inició de manera sistemática la guerra informativa interna y externa, ante las denuncias masivas de fraude electoral y el estallido de protestas en las principales ciudades del país. La reacción gubernamental no solo se centró en la represión de las manifestaciones, sino que se desplegaron rápidamente campañas coordinadas en el espacio digital: se crearon ejércitos de trolls y *hackers* con el objetivo de controlar la narrativa *online*, inundando foros y redes sociales con mensajes favorables al gobierno y promoviendo teorías de injerencia occidental. Entre ellas se encuentra el nacimiento de comunidades de trolls en redes propias del país como VKontakte y Odnoklassniki, que se emplearon para monitorear, desacreditar y distraer la discusión pública nacional e internacional sobre el fraude y las protestas (Milosevich-Juaristi, 2021).

VKontakte (VK), que surgió inicialmente como una imitación del occidental Facebook, se consolidó rápidamente como la red social dominante en Rusia. A pesar de su gran popularidad, la plataforma mantenía una dualidad, siendo un centro de comunicación masiva a la vez que un conocido refugio para la piratería de contenidos sin licencia. Este estatus cambió drásticamente en 2014 con la toma de control por parte del Kremlin. El fundador, Pavel Durov, vendió su participación y huyó del país en medio de una falsa acusación de atropello, articulando su desilusión con la frase: "lo que posees, tarde o temprano, te posee" (Colon, 2025). La plataforma fue adquirida por el grupo mail.ru, propiedad de Alisher Usmanov, un oligarca con fuertes vínculos con Vladimir Putin, lo que aseguró la capacidad del Estado para ejercer presión, censura y vigilancia. Este control permitió a la agencia Roskomnadzor solicitar y obtener el bloqueo del acceso a comunidades nacionalistas ucranianas y a grupos que mencionaban al opositor Alexey Navalny (Soldatov, & Borogan, 2015). Dada la popularidad de

VKontakte en Ucrania, con más de 20 millones de usuarios, la plataforma se convirtió en un vector de gran importancia para la guerra de la información rusa. En respuesta a esta penetración propagandística, Ucrania tomaría la medida drástica de prohibir VKontakte y otras redes sociales rusas en 2017 (Vilmer, & et al., 2018). Durov, por su parte, crearía la red Telegram, que sería famosa por su uso por parte de grupos de todo tipo, sin facilitar información a cualquier autoridad, hasta su detención en París en 2024 por 12 delitos relacionados con transacciones ilícitas, pornografía infantil, fraude y negativa a comunicar información a las autoridades (El Pais, 2024).

En 2013 se publica la llamada doctrina Gerasimov, que es considerada por muchos el enfoque ruso contemporáneo sobre la guerra híbrida, articulado por el general Valeri Gerasimov, Jefe del Estado Mayor de las Fuerzas Armadas de Rusia (Герасимов, 2022). Propone que los conflictos armados modernos combinan métodos militares convencionales con una amplia gama de acciones no militares, como la desinformación, presión económica, subversión política, operaciones psicológicas y cibernéticas. A diferencia de la concepción occidental de la guerra híbrida, Gerasimov interpreta "métodos híbridos" como aquellos que sobrepasan la intervención militar directa, dando centralidad a la manipulación informativa y la desestabilización interna mediante el apoyo de actores internos en el país objetivo, el uso de revoluciones, propaganda mediática, sabotaje, operaciones encubiertas y campañas de falsificación de hechos. Su hipótesis central es que cuantas más acciones indirectas se lleven a cabo sobre los adversarios, menos necesario será el uso de la fuerza militar directa. Esta doctrina diluye la frontera entre la paz y la guerra, permitiendo operar en una "zona gris" donde el Estado podría agredir y alcanzar objetivos estratégicos sin recurrir a invasiones convencionales, siendo la supremacía informativa y comunicacional el elemento esencial para el éxito. Algún experto, como Galeotti, difiere de que esto sea una doctrina nueva, sino más bien la misma usada desde hace mucho tiempo (Galeotti, 2019).

En 2013 se fundaría el *Internet Research Agency* (IRA) en San Petersburgo, lo que sería conocida como una de las más influyentes "fábrica de trolls". Este organismo fue uno de los pioneros en coordinar miles de perfiles falsos en redes como Twitter, Facebook y VKontakte, extendiendo rápidamente su acción igualmente fuera de Rusia, con el fin de intervenir y polarizar tanto en el debate interno como en escenarios internacionales. La periodista finlandesa Jessikka Aro se dedicó a diseccionar las técnicas de intimidación utilizadas por el IRA, financiada y secretamente establecida por el oligarca Yevgeny Prigozhin, un persona conocida como el "Chef de Putin". Esta organización operaba con cientos de empleados en

departamentos altamente compartimentados, dedicados a la creación de sitios web falsos, memes y contenido viral, utilizando innumerables cuentas falsas (*sock puppets*) en redes sociales y foros de diarios digitales occidentales, con el objetivo de desorganizar a la oposición interna y extender las operaciones de influencia al extranjero (Aro, 2020). El perfil de 818 de sus trabajadores reflejaba principalmente periodistas y de publicidad (39,18%), economistas y marketing (10,82%), ingeniería (8,54%) y artes (8,38%), con una media de 27 años y unos 1.000$ de salario medio mensual (Poliakoff, 2025). Lamentablemente, el costo personal para Aro fue muy elevado, pues se convirtió en víctima de las mismas tácticas que buscaba denunciar (fue presentada en redes como drogadicta o desequilibrada mental), llegando incluso a recibir acoso en la vida real de grupos criminales tanto neonazis como de extrema izquierda (Blanco, 2017). En 2019, Aro recibió el premio *International Women of Courage* por parte del Departamento de Estado de EE.UU., pero le fue retirado justo antes de la ceremonia por criticar en redes al presidente Donald Trump (Sánchez, 2019).

La campaña de desinformación durante la anexión de Crimea (Ucrania) en 2014 marcó un salto estratégico, ya que se desplegaron campañas masivas de saturación mediante la inundación de contenidos, que hicieron convivir noticias falsas, emociones manipuladas y relatos elaborados por bots y medios estatales (Dawson, & Innes, 2021). En ese contexto, la caída del vuelo MH17 (vuelo comercial de Malaysia Airlines que fue derribado el 17 de julio de 2014 mientras sobrevolaba el este de Ucrania, mientras iba de Ámsterdam a Kuala Lumpur con 298 personas. Tras investigaciones, se determinó que fue alcanzado por un misil de fabricación rusa lanzado desde una zona controlada por separatistas pro-Kremlin) se utilizó como caso a gran escala amplificando la versión oficial rusa y desacreditando a periodistas y fuentes independientes mediante ataques coordinados y manipulaciones en tiempo real, usando *trending topics* y comunidades simuladas de apoyo pro-Kremlin (RTVE, 2020; Espaliú-Berdud, 2023). Las campañas en Crimea en 2014 son el molde de la guerra híbrida moderna, en las que narrativas coordinadas sobre el supuesto peligro de los protestantes en el Maidan en Kiev, la protección de la minoría rusa y la legitimidad del referéndum se impulsaron en redes sociales mediante cuentas falsas, bots y medios estatales (Galeotti, 2020). Imágenes distorsionadas y testimonios creados lograron manipular emocionalmente a la audiencia, con ataques informativos directos a la infraestructura ucraniana y sabotaje digital que dificultó el acceso a información. La manipulación del debate público incluía la administración de grupos falsos en Facebook y foros, la creación de *trending topics* y la simulación de movimientos ciudadanos, logrando

viralizar posiciones rusas y desacreditar al periodismo internacional (Soldatov, & Borogan, 2015).

Entre 2014 y 2020, la campaña "Infección Secundaria" (*Secondary Infektion*) se consolidó como uno de los ejemplos más sofisticados de desinformación masiva a escala internacional. Durante seis años, agentes vinculados al Kremlin distribuyeron más de 2.500 contenidos falsos y documentos manipulados en siete idiomas diferentes, empleando alrededor de 300 plataformas *online* (Colon, 2025). El objetivo principal era crear y amplificar narrativas polarizantes, atacar adversarios políticos y sembrar confusión, especialmente en Europa y América. Estas campañas influenciaron debates políticos, procesos electorales y situaciones de conflicto, como el caso ucraniano y la difusión de bulos sobre el origen de enfermedades, utilizando falsificación de documentos y su viralización mediante cuentas falsas. Para aprender cómo hacer campañas efectivas, el IRA envió a dos mujeres en 2014 a Estados Unidos durante tres semanas para recopilar inteligencia y monitorear grupos sociales y detectar temas que afectaran a la sociedad (Jancowicz, 2020). En 2015 se hicieron pruebas de reparto gratuito de perritos calientes en Times Square de Nueva York y seguimiento de cuánta gente se acercaba mediante webcams públicas para comprobar la eficacia de distintas estrategias de campañas (Colon, 2025).

Entre 2016 y 2024, la sofisticación técnica y el alcance de las operaciones rusas se incrementaron notablemente. Casos emblemáticos como *Ghostwriter* (publicación de mensajes y vídeos falsificados para hacer perder la confianza en instituciones de Polonia, Ucrania y Alemania) (del Castillo, 2022; Sanger, Barnes, & Conger, 2022), la interferencia en las elecciones presidenciales estadounidenses de 2016 (en las que el presidente americano saliente Barack Obama instó al presidente ruso Vladimir Putin a "detener eso" en la reunión del G20 en Hangzhou, China) (Colon, 2025), el escándalo *MacronLeaks* en Francia (2017) y las campañas antivacunas ilustran cómo se adoptaron nuevos métodos: manipulación masiva de comentarios en medios occidentales y redes, creación de narrativas alternativas, empleo de vídeos y audios "*deepfake*", y saturación informativa para polarizar y desestabilizar sociedades occidentales (Milosevich-Juaristi, 2020). Hay que tener presente que existen estudios que muestran que hubo un 15% más de mortalidad entre votantes republicanos en Estados Unidos que entre demócratas por una menor vacunación en época de pandemia por covid-19 (Wallace, Goldsmith-Pinkham, & Schwartz, 2023).

El aparato estatal ruso se apoyaría en estructuras complejas como SORM y Roskomnadzor, que refinan el control sobre el tráfico digital y la censura. SORM permite la interceptación masiva de comunicaciones, facilitando la vigilancia y

la represión de toda disidencia, mientras que Roskomnadzor actúa bloqueando webs enteras y exige a las tecnológicas que almacenen datos en Rusia, lo que les dota de poder para perseguir y silenciar voces independientes. El efecto combinado de estos sistemas es doble, pues dificulta el acceso a información alternativa y blinda la propaganda oficial (Quénel, 2023).

La realidad del IRA muestra que, ante las sanciones y el endurecimiento de las políticas tecnológicas occidentales, sus métodos se han descentralizado pues hoy existen micro-agencias, empresas pantalla y subcontratistas que, lejos de San Petersburgo, trasladan el trabajo de manipulación hacia la automatización avanzada, el uso de inteligencia artificial y la producción de *deepfakes*. La inundación del espacio informativo ("*littering the information space*") se intensifica con campañas en África, América Latina y Asia, donde la colaboración con el Grupo Wagner y otras entidades permitió crear portales que simulan medios locales y amplificar narrativas prorrusas insertadas en cada contexto sociopolítico (European External Action Service-EEAS, 2025).

La flexibilidad de la industria se evidencia en el ajuste temático y geográfico, así en los últimos años la desinformación rusa ha priorizado campañas sobre la pandemia, la guerra en Ucrania, la crisis energética europea y temas de polarización social, migrando a plataformas emergentes como Telegram y TikTok para sortear bloqueos. El modelo de manguera de falsedades ("*firehose of falsehoods*" mediante la saturación y atomización de mensajes) ha conseguido elevar el ruido digital y hacer cada vez más difícil la moderación. El sistema ruso mantiene su capacidad de adaptación, integración con el aparato estatal y un proceso permanente de innovación digital, lo que plantea unos desafíos extraordinarios.

A raíz del ciberataque estadounidense contra la Agencia Rusa de Investigación de Internet (IRA) antes de las elecciones de medio término en EE. UU. de 2018 (Sanger, Barnes, & Goldman, 2020), Rusia aceleró la transformación de su ecosistema digital y reforzó sus medidas de protección estatal, llevando a una reconfiguración operativa notable del IRA y sus granjas de trolls. Por ello implementó prácticas avanzadas de anonimato como el uso intensivo de VPNs, identidades falsas y servidores fuera de Rusia y Europa, ampliando su campo de acción desde redes sociales tradicionales hacia plataformas emergentes como Reddit, Tumblr y ampliando hacia canales en otros idiomas para audiencias latinoamericanas y nuevos mercados vulnerables (University of New South Wales, 2023).

En paralelo, tras el ciberataque el Kremlin puso en marcha el proyecto RuNet, una infraestructura nacional de internet creada tras la Ley de Internet Soberana de 2019. Su finalidad era doble, pues por un lado quería ofrecer resiliencia frente a ciberataques y aislamiento internacional, y por otro, garantizar un

control absoluto del tráfico digital desde organismos como Roskomnadzor y el FSB (Servicio Federal de Seguridad de la Federación Rusa, el considerado heredero del antiguo KGB), que pueden inspeccionar, bloquear, redirigir o filtrar cualquier contenido sin supervisión judicial. RuNet se basa en la instalación obligatoria de equipos de inspección profunda de paquetes (DPI) y sistemas de DNS y certificados SSL nacionales, lo que permite al Estado mantener el funcionamiento de servicios y páginas rusas incluso en caso de desconexión global (Epifanova, 2020). En 2023, tras el motín de Prigozhin que desafió al presidente Putin, miembros del Servicio Federal de Seguridad ruso tomaron el control del grupo *Patriot Media* y empezaron a desmantelar su estructura, cerrando el 30 de junio de dicho año. A partir de ese momento muchas de las cuentas en redes sociales del IRA, comenzaron a llamar traidor a su exjefe, justo al que alababan un mes antes, y participar en otro tipo de campañas, dando pie a deducir que se tomó su control a partir de entonces (Sauer, 2023).

La consolidación de RuNet ha tenido un impacto profundo en la sociedad rusa, facilitando la censura, el bloqueo de medios independientes y la marginación de narrativas alternativas (Flashpoint Intel Team, 2023). Las autoridades han aumentado la presión sobre redes sociales internacionales y servicios VPN, restringiendo el acceso a otras fuentes informativas y persiguiendo judicialmente a periodistas y comunicadores críticos. Este monopolio informativo estatal, reforzado por la capacidad técnica de interceptar y analizar comunicaciones, ha reducido drásticamente la pluralidad mediática y la libertad de prensa, empujando a muchos ciudadanos y profesionales hacia el exilio digital. Tras una prueba en 2023 de unas horas (del Castillo, 2023), en noviembre de 2025 se anunció el control absoluto de la red a partir del 1 de enero de 2026, dando legalidad a cualquier bloqueo de la red y el acceso a contenido no permitido (Cuesta, 2025).

En abril de 2024 se celebró en San Petersburgo una reunión de responsables de seguridad a la que asistieron personas de 40 países, en la que se ofreció una alianza de protección mutua contra la influencia occidental bajo el concepto de protección de valores tradicionales. En dicha propuesta se ofreció ciberseguridad y control de redes sociales para controlar el espacio informativo mediante tecnologías avanzadas de empresas rusas (Greene, Soldatov, & Borogan, 2024).

1.5. LA GUERRA DE UCRANIA, TODO SE ACELERA

La invasión rusa de Ucrania en 2022 ha consolidado el conflicto como una guerra híbrida de alta intensidad donde el frente militar y el informativo-cognitivo avanzan en paralelo, en la llamada "guerra o confrontación de la información"

(*informatsionnaya voina*) (de Pedro, & et al., 2023). La red social Telegram emergió como el principal campo de batalla digital después de que el Kremlin bloquease plataformas occidentales como Facebook y Twitter/X, convirtiendo la app en el vector central para narrativas prorrusas y en herramienta de uso militar, tanto para propaganda como para comunicaciones entre unidades rusas ante las deficiencias de sus propios sistemas (Buziashvili, & et al., 2024).

En este entorno, Rusia desplegó campañas sofisticadas como *Doppelgänger* (RRN) (vocablo para denominar al doble fantasmagórico de una persona en la mitología alemana) (Quénel, 2023), que desde 2022 ha clonado centenares de dominios de medios de comunicación e instituciones europeas para difundir contenidos falsos amplificados con cuentas automatizadas y anuncios de pago, orientados a negar crímenes de guerra como Bucha, acusar a Ucrania de nazismo, deslegitimar las sanciones y modular la opinión pública en Europa y el Sur Global, mientras recurrían también a *deepfakes* como el vídeo de un Zelenski falso llamando a rendirse y mantenía activa su red de trolls heredera de la IRA bajo un marco de férrea censura interna y penas de hasta 15 años por "mentir sobre el ejército" (Colon, 2025). Igualmente se creó la granja de trolls *CyberFront Z* en San Petersburgo, para difusión de teorías conspirativas, fomento de dudas y división principalmente hacia países occidentales en diferentes redes sociales mediante unos 100 empleados escritores por turno, además de voluntarios, por un salario mensual de unos 431,96$ (Gilbert, 2022).

Frente a ello, Ucrania convirtió el dominio digital en un eje central de resistencia mediante una combinación de humor, movilización ciudadana y liderazgo comunicativo. La guerra de memes ucraniana utilizó cuentas oficiales y no oficiales para ridiculizar a Putin y a los invasores, mantener la atención internacional y reforzar la moral interna, al tiempo que Zelenski se proyectaba en Instagram como figura de una nación unida que resiste, contrarrestando rumores de huida. Paralelamente, estructuras como el "I-ejército" (Iriarte, 2025) canalizaron la participación de voluntarios digitales en tareas de ciberdefensa, OSINT y denuncia, incluyendo tácticas como el uso de perfiles falsos en apps de citas y el rastreo de fotos en Telegram para geolocalizar posiciones rusas, lo que permitió incluso ataques de precisión contra instalaciones de Wagner. Las grandes plataformas tecnológicas occidentales abandonaron en buena medida la neutralidad, por ejemplo Meta desmanteló redes rusas y flexibilizó temporalmente sus normas de discurso contra militares rusos, mientras Google y Microsoft bloquearon campañas de influencia y ayudaron a asegurar infraestructuras digitales ucranianas. En contraste, TikTok se consolidó como canal clave donde

medios estatales chinos amplificaron el marco ruso de "operación militar especial", sin mostrar un esfuerzo equivalente para limitar la propaganda financiada por el Kremlin (Colon, 2025).

Un caso especial a destacar es el de NAFO (*North Atlantic Fellas Organization* - Organización de Compañeros del Atlántico Norte), un movimiento digital descentralizado nacido tras la invasión rusa de Ucrania en 2022, fundado en mayo de ese año por el artista polaco Kamil Dyszewski ("Kama") como iniciativa para recaudar fondos para la Legión Georgiana y otras unidades ucranianas (Munk, 2025). Su símbolo es un perro Shiba Inu, derivado del meme "Doge", vestido con uniformes y equipamiento militar y a menudo superpuesto a una versión modificada del logotipo de la OTAN, que funciona como avatar distintivo de sus miembros ("*fellas*"). A cambio de donaciones, los simpatizantes reciben versiones personalizadas de este personaje, lo que ha permitido canalizar importantes sumas hacia material, drones y otros recursos para el ejército ucraniano, mientras el grupo se consolida como comunidad transnacional de apoyo a Kiev y contra la propaganda del Kremlin (Mejova, & et al., 2025).

En redes sociales, NAFO actúa como ciberguerrilla de memes de cuentas que usan el humor, troleo inverso y respuesta coordinada para ridiculizar a portavoces pro-Kremlin, inundar hilos con memes y desviar la conversación frente a narrativas de Moscú, a la vez que señala crímenes de guerra o fracasos militares rusos (Wywiał, 2023). Su estructura es informal, sin jerarquía ni código deontológico claros, lo que favorece la creatividad pero también ha generado críticas cuando algunos miembros atacan a analistas occidentales simplemente por describir dificultades ucranianas (Kasianenko, & Boichak, 2024). El ecosistema ruso los presenta como un proyecto de guerra informativa de la CIA o el Pentágono, mientras portavoces como María Zajárova los acusan de encarnar odio y xenofobia. En contraste, NAFO ha sido legitimado simbólicamente por dirigentes como Ben Wallace, Kaja Kallas u Oleksii Reznikov, que se han declarado "miembros honoríficos" y han participado en la primera "cumbre NAFO" en Vilna en 2023, consolidando al grupo como actor visible en la guerra de la información (Iriarte, 2025).

La cantidad de mensajes de difusión de desinformación y/o odio se incrementa ante el aumento de medios, actores, bajos costes y facilidad de equipos de manera exponencial. En 2022 la empresa Meta revela que sólo en el primer cuatrimestre de 2022, de enero a marzo, eliminó en Facebook 21,7 millones de mensajes que eran de odio y violencia, cuando el cuatrimestre anterior habían sido 12,4 millones (Rosen, 2022). Por unos u otros motivos la empresa eliminó 1,6 mil millones de cuentas falsas en ese mismo periodo y calculó que al menos el 5% de

todos los usuarios activos en Facebook eran falsos, con un gran incremento del odio organizado tanto en esa red como en Instagram (Hutchinson, 2022).

El análisis de emociones y sentimientos de los discursos en redes sociales comienza a seguirse de manera masiva, como muestra en 2023 la empresa australiana *Fivecast* que se llevó el concurso por 20 millones de dólares del Gobierno de Estados Unidos para la Oficina de Aduanas y Protección Fronteriza (CBP), que es parte del Departamento de Seguridad Nacional de dicho país, para captar, por ejemplo, ira, disgusto, miedo o sorpresa (aspectos que se explicarán en la segunda parte de este libro). Sus análisis mediante técnicas de aprendizaje automático no sólo recopilan datos públicos de internet, sino de todo lo expuesto en redes sociales e incluso la *dark web* (zona de internet que no aparece en buscadores tradicionales ni se puede ver con navegadores web comunes, y que a menudo se asocia con uso de actividades ilegales o con lugares donde la censura es común) (Escribano, 2023). Se llega incluso en varias investigaciones periodísticas a explicar, en ese mismo año, cómo las autoridades de inmigración de EE.UU. crearon y utilizaron perfiles falsos en redes sociales para investigar a las personas que solicitaban ayudas de inmigración (Bhuiyan, & Levin, 2023).

En 2022 el área de Ciencia y Tecnología de la OTAN reconoce como estratégica la guerra cognitiva, creando en abril de 2025 20 grupos de trabajo relacionados con la investigación en este campo, varios de ellos centrados en los medios, redes sociales y demás operaciones de información (Blatny, & Søndergaard, 2025).

La desinformación crece y crece, así un informe del *Ministère de L'Europe et des Affaires Étrangères* francés de 2025, da varios datos muy significativos. Primero, que el gasto medio de Rusia en operaciones de manipulación de la información fue de 1.600 millones de dólares en 2023, con destino hacia 90 países y más de 500 incidentes registrados en redes sociales y otros canales de internet, 42 de ellos contra elecciones europeas. El estudio de estas operaciones ofrece cuáles son sus primeros 10 países objetivo (Ministère de L'Europe et des Affaires Étrangères, 2025): Ucrania (37,4%), Francia (21,2%), Alemania (8,1%), Mali (7,6%), Niger (7,6%), Estados Unidos (4,5%), Burkina Faso (4,5%), España (4%), Polonia (3,5%) y Bélgica (3,0%). Un reportaje periodístico del canal franco-alemán Arte.tv expuso que el grupo de Prigozhin gastó en campañas de desinformación 1.676.970€ únicamente en el mes de noviembre de 2019 (Jousset, 2022). Cabe señalar, para comparar, que la unidad que EE.UU. tenía contra la lucha de la desinformación extranjera en redes sociales tenía un gasto de 61 millones de dólares al año y unos 120 empleados en 2024, hasta que fue cerrada por la administración Trump en 2025 con el argumento de que sus actividades se habían

convertido en un mecanismo de censura y de violación de la libertad de expresión (Psaledakis, 2025). Mientras, Europa movilizó 50 millones contra la desinformación entre 2015 y 2020 según el Tribunal de Cuentas Europeo (European Court of Auditors, 2021), mientras que el presupuesto en su división de lucha contra desinformación del Servicio Europeo de Acción Exterior creció hasta 14,6 millones de € en 2023 (European External Action Service, 2022), mismo presupuesto consolidado para 2024 (European External Action Service, 2023).

Noruega, al convertirse tras las sanciones a Rusia por la guerra en Ucrania en el mayor suministro de petróleo de la Unión Europea, ha recibido un gran incremento de ataques de desinformación. Según el Proyecto *American Sunlight* y *Bellona*, una organización ambiental noruega, han detectado 697 cuentas en X organizadas por una entidad rusa con spam masivo a organismos oficiales, ONG e instituciones educativas de que la industria petrolífera noruega crea problemas medioambientales, desastres naturales y trabaja con criptomonedas para desviar impuestos. La campaña, denominada *EcoBoost*, descubierta en febrero de 2025, hasta el 31 de julio de ese mismo año había publicado 602.668 mensajes coordinados, atacando por un lado a las empresas petrolíferas, difusión de teorías conspiranoicas y por otro ridiculizando a los movimientos ecologistas, además de intentar convertirlos en activos ignorantes (*useful idiots*) (Shultz, 2025).

En 2024 el investigador Owen Jones, referencia en el campo de análisis de la desinformación algorítmica, expuso en su informe "*The Qatar Plot*" (Dsouza, & Jones, 2024) campañas de alcance mundial que promueven principalmente ideas de extrema derecha, conspiraciones, inmigración (con la idea del "Gran Reemplazo", en el que se promueve que élites quieren reemplazar a europeos blancos por personas negras o morenas, particularmente musulmanes), y ataques al mundo musulmán, con un alcance de mínimo de 50 millones de personas en distintas redes sociales, apareciendo incluso anuncios en las pantallas de *Times Square* de Nueva York. Sus análisis detectaron estas campañas actuando sobre Francia, Alemania, Suecia, España, Malta, Croacia, Reino Unido y Estados Unidos, mientras que en Líbano y Arabia Saudí fomentaban tensiones entre distintas partes de la sociedad. Las cuentas ocultaban sus identidades a través de *proxies* en Vietnam (Jones, 2024).

Entre 2024/25 las elecciones a Rumania tuvieron que ser repetidas por injerencias en la campaña. En la primera vuelta del 24 de noviembre de 2024, el candidato ultraderechista Calin Georgescu, que apenas tenía un 1% de intención de voto semanas antes, resultó ser el candidato más votado con un 22,9%, basando su campaña en el desencanto económico, echando la culpa a la Unión Europea y la OTAN (a los que denominaba élite globalista satánica y pedófila) por la guerra

en Ucrania, renunciar a la "basura" del feminismo frente a la belleza natural de las mujeres, y la promesa de restablecer relaciones con Rusia (Bârgăoanu, 2025). Su perfil personal incluía la admiración por figuras fascistas históricas, posturas antivacunas, escepticismo climático y la difusión de teorías de la conspiración (como que el hombre no llegó a la Luna o la existencia de "no humanos" en la ONU). Se encontró que la red *Portal Kombat*, una red de 25.000 cuentas en TikTok y varios canales de Telegram apoyaron a dicho candidato mediante masivas campañas astroturfing (Secrétariat Général de la Défense et de la Sécurité Nationale-SGDSN, 2025a), además de 85.000 ataques informáticos. Tras encontrar donaciones sospechosas por más de un millón de euros a través de TikTok a dicho candidato, el Tribunal Constitucional del país anuló la primera vuelta (Olari, 2025) y detuvo a un mercenario con grandes cantidades de dinero y armas destinadas a crear disturbios para la segunda vuelta (Iriarte, 2025).

Estados Unidos denegó en diciembre de 2025 el visado a cinco europeos que trabajan en la regulación de plataformas digitales y en la lucha contra el discurso de odio y la desinformación, acusándolos de "coaccionar" a empresas tecnológicas estadounidenses para "censurar" puntos de vista protegidos por la Primera Enmienda. Entre los afectados estaban el excomisario europeo Thierry Breton, impulsor de la Ley de Servicios Digitales (DSA), Imran Ahmed (*Centre for Countering Digital Hate*), Clare Melford (*Global Disinformation Index*) y las activistas alemanas Anna-Lena von Hodenberg y Josephine Ballon (*HateAid*). La administración Trump justificó la decisión como una defensa de la libertad de expresión estadounidense frente a normas europeas que exigirían a las plataformas moderar contenidos de odio, desinformación y abusos en línea. Por su parte Gobiernos europeos, organizaciones de derechos humanos y algunas entidades de EE.UU. denunciaron estas restricciones como un ataque político a aquellos que combaten el odio y la desinformación, y una injerencia en la soberanía europea, poniendo la retórica de la libertad de expresión como una forma de frenar la rendición de cuentas de las empresas del entorno digital (France24, 2025).

1.6. LA DESINFORMACIÓN EMPRESARIAL

La industria de la desinformación se ha consolidado como un negocio global que va mucho más allá de las operaciones estatales y motivos geopolíticos, articulado en torno a la "desinformación como servicio" (campañas por encargo) y la "desinformación como producto" (contenidos falsos monetizados por publicidad). Un estudio de Roberto Cavazos para la Universidad de Baltimore y CHEQ estimó que el coste anual de pérdidas que causa la desinformación rondó los

78.000 millones de dólares en 2019 (CESIE, 2022), incluyendo unos 39.000 millones en pérdidas bursátiles, 9.540 millones en gasto reputacional corporativo y alrededor de 9.000 millones ligados a desinformación sanitaria (Maurice, 2024). Estas pérdidas están ligadas principalmente a la pérdida de reputación que hace bajar la confianza de los consumidores en la marca y sus productos (di Domenico, & Ding, 2023). En paralelo, *NewsGuard* y *Comscore* calcularon que los sitios de desinformación ingresaron unos 2.600 millones de dólares al año en 2021 en publicidad programática, de los que 1.620 millones de recaudación procederían del mercado estadounidense. Para contextualizar esta cifra, por cada 2,16 dólares que la publicidad digital envía a periódicos legítimos en EE.UU., los anunciantes pagarían un dólar más a sitios de desinformación (Skibinski, n.d.).

En el ámbito laboral, la desinformación y la vigilancia se han integrado en servicios comerciales que monitorizan sistemáticamente la actividad digital de empleados, bajo rótulos como "*social media monitoring*" (monitor de redes sociales) o "*employee risk monitoring*" (monitor de riesgo de empleados). Estas herramientas rastrean publicaciones, comentarios y cambios de perfil en redes sociales para detectar supuestas "fugas de información", críticas internas o indicios de "riesgo reputacional", extendiéndose incluso a redes privadas y seguimiento de estados de baja médica en algunos contextos, pese a las restricciones legales en países europeos. A ello se suma la industria del *union-busting* (antisindicatos), que mueve unos 340 millones de dólares anuales en EE.UU. y que identifica y sigue líderes sindicales, segmenta colectivos internos y lanza mensajes que siembran desconfianza de los trabajadores hacia sus representantes, a menudo usando medias verdades o insinuaciones ("os van a quitar beneficios", "esta gente tiene otra agenda", etc.) (Walicek, 2023).

También se emplean sistemas de analítica interna (correo, chat corporativo, herramientas colaborativas) para mapear redes de influencia dentro de la empresa: quién habla con quién, quién es nodo central, quién muestra descontento. Algunas consultoras venden estos servicios como forma de "gestión de riesgos internos", donde identifican "potenciales agitadores", "líderes informales" o "posibles denunciantes" y recomiendan estrategias de neutralización (moverlos de equipo, aislarlos comunicativamente, desacreditar sus mensajes, etc.) (Mettler, 2024; Lipscomb, 2025).

En el extremo más agresivo operan firmas de *dark* o *black PR* (propaganda negra o negativa) que venden campañas de difamación en redes, reseñas falsas, intoxicación de resultados de búsqueda y creación de cuentas anónimas para acusar de incompetencia o mala conducta, técnicas aplicables tanto contra

competidores como contra exempleados o denunciantes (Rodríguez Fernández, 2021; Ennis, 2023).

El mercado de desinformación como servicio es, además, sorprendentemente barato, según investigaciones de *Recorded Future* existen tarifas para hacer campañas contra la competencia aproximadamente de entre 3.000 a 7.000 dólares (con un equipo de un coordinador y 5 a 10 trabajadores haciendo campaña en redes) para un ámbito de mercado local o de nicho (Insikt Group, 2019). Para campañas a nivel de país (agencia con entre 20 a 50 trabajadores dedicados) ronda de 25.000 a 80.000 dólares (Paulo, 2022), mientras que una internacional con posible impacto bursátil y regulatorio de 250.000 a 1 millón de dólares. Estos precios varían también dependiendo del área geográfica, siendo las empresas indias y sudeste asiático las que están en horquillas de tarifas más bajas frente a las occidentales. Entre ellas, Filipinas e Indonesia destacan en número de empresas que realizan este tipo de operaciones de alquiler de desinformación (Hapal, 2024).

Según el informe *Industrialized Disinformation* del *Oxford Internet Institute*, desde 2009 a 2018 se habrían gastado al menos 60 millones de dólares en contratar firmas privadas de propaganda computacional, con más de 65 empresas operando en 48 países, aunque los autores subrayan que las cifras reales probablemente son muy superiores (Bradshaw, Bailey, & Howard, 2020).

1.7. CASOS POR PAÍSES O ÁREAS

1.7.1. Italia

Italia se considera un laboratorio donde se han llevado a cabo muchos experimentos políticos, y se convirtió en el "Silicon Valley del populismo", anticipándose en muchos años a lo que sucedería en otros lugares, según uno de los autores que mejor describen este país, Giuliano da Empoli (2020). De esta manera Italia constituye un buen ejemplo de cómo las dinámicas digitales pueden transformar el panorama político mediante estrategias orientadas a explotar las emociones, segmentar audiencias y manipular percepciones colectivas en favor de intereses partidistas.

A inicios de los 2000, el surgimiento del Movimiento 5 Estrellas (M5S) marcó un punto de inflexión. Gianroberto Casaleggio, experto en comunicación digital, trasladó sus ideas al blog del cómico Beppe Grillo, convirtiéndolo en espacio de crítica al *establishment* y de propuestas populares más emocionales que racionales. Desde 2012, la proliferación de sitios afines generó multitud de

"realidades paralelas" con titulares provocadores y desinformativos que reforzaban la indignación como motor político. El control sobre los relatos mediáticos se intensificó con la sección "periodista del día" en el blog del M5S, donde se exponía públicamente a periodistas críticos, presentados casi como símbolos de una prensa corrupta, contribuyendo a una atmósfera que facilitaba el hostigamiento digital. Este clima, según Reporteros Sin Fronteras, contribuyó significativamente a las restricciones sobre la libertad de prensa en Italia (Reporteros Sin Fronteras, 2017). El partido M5S llegaría en 2013 a ser el tercer partido en votos, y el de mayor representación en el parlamento en 2018.

Poco después, la estrategia de aprovechamiento digital fue replicada y refinada por la Lega de Salvini, que en 2014 creó un sofisticado sistema de análisis de redes sociales llamado "la Bestia". Este dispositivo digital monitorizaba sistemáticamente las respuestas a mensajes y publicaciones del líder nacionalista, trabajaba los datos y elaboraba toda esa información en eslóganes y campañas personalizadas, con el objetivo de optimizar el impacto mediático y electoral. La recogida de datos llegó a desplegarse incluso en entornos lúdicos, como el juego *online* "Vinci Salvini". Otro actor que recogería este modelo del M5S de forma reconocida sería el británico Nigel Farage, del partido UKIP famoso por estar al frente de la campaña del *Brexit* para la salida del Reino Unido de la Unión Europea, pero con el tiempo tomarían caminos distintos (The Guardian, 2017).

Estos movimientos se consolidaron en un contexto de ataque recurrente a la prensa y de utilización estratégica de temas socialmente sensibles, siendo la inmigración uno de los ejes centrales de la desinformación electoral en el periodo 2017-2018. Numerosos estudios identificaron la proliferación de sitios web mixtos (información y titulares falsos o sesgados que coincidían con otros en Brasil) y la amplificación sistemática de estos en plataformas como Facebook y Twitter (Applebaum, 2020). Las investigaciones mostraron que en 2019, la exposición a desinformación en plataformas de Meta en italiano (al igual que en español) era mucho mayor que en inglés (68% y 70% respectivamente frente al 29% en inglés), destapando carencias en los sistemas de advertencia (Avaaz, 2020).

En los últimos años, el panorama ha evolucionado hacia una internacionalización del riesgo, con la aparición de campañas de desinformación pro-Kremlin y una creciente sofisticación de las técnicas empleadas por actores nacionales y extranjeros (como por ejemplo la evolución de canales antivacunas en pandemia Covid-19 que derivaron con el tiempo a contenido pro-Kremlin y anti-Otan con la guerra de Ucrania) (Institute for Strategic Dialogue, 2025). Una investigación propia, estudiando las reacciones en X/Twitter tras la dimisión del presidente Draghi en 2022, pudo comprobar, además de la alta polarización,

conspiraciones y empleo de discurso de odio, la aparición de numerosos mensajes aparentemente de italianos, que en verdad procederían desde Albania y Tailandia (Mottareale-Calvanese, Arce-García, & Said-Hung, 2025).

1.7.2. México

México se ha erigido como otro laboratorio internacional para la industria de la desinformación digital, con raíces que se remontan a las elecciones de 2006, cuando empezaron a emplearse tácticas para atacar blogs y foros de oposición política tras disputadas contiendas electorales. El verdadero salto tecnológico llega en 2009 con la expansión de Twitter durante la pandemia del virus AH1N1, donde periodistas como Alberto Escorcia documentan los primeros ensayos masivos de manipulación informativa en redes sociales (Iriarte, 2024). El punto de inflexión lo marca la campaña de Enrique Peña Nieto en 2012, donde se documenta por primera vez el uso sistemático de granjas de bots y trolls, con operaciones que incluían ejércitos digitales contratados (algunos con presupuestos superiores a 600.000 dólares y más de 30.000 cuentas automatizadas) para posicionar *hashtags* influyentes como #EsMomentodeMéxico (Peinado, Palomo, & Galán, 2018).

La estructura operativa de esta industria combina actores estatales, partidos políticos y contratistas privados, configurando equipos de trabajo con personal dedicado y financiamiento constante, con empresas que compiten ofreciendo paquetes integrales, con acceso a decenas de miles de bots y pagos mensuales que superan el millón de dólares. Un ejemplo algorítmico sucedido en México sería el caso de la "Secta 100tifika", colectivo que desde Málaga, España, fue responsable de campañas de manipulación y provocación de pánico en saqueos de 2017, logrando viralizar *hashtags* como #SaqueaUnWalmart con apenas 485 cuentas y más de 1.500 mensajes coordinados (24 Horas Puebla, 2017).

Distintos informes, como el del *Oxford Internet Institute* y laboratorios como Signa Lab de ITESO, coincidieron en que entre el 18% y el 27% del contenido político en Twitter mexicano podría estar generado por bots. En 2018 se documenta que el partido político PRI, así como sus partidos oponentes, recibieron oferta de servicios por parte de Cambridge Analytica (Semple, & Ahmed, 2018; Kaiser, 2019), aunque un directivo de esta empresa fue grabado afirmando que operaron en el país (El País, 2018). En 2020 se detectó una oficina con 7 empleados del IRA del ruso Prigozhin en México, para difusión de campañas en redes en español (Jousset, 2022).

En las elecciones de 2024 se denunció el despliegue de más de 87.000 cuentas bot que generaron un millón de mensajes contra Claudia Sheinbaum, con

una inversión de 20 millones de dólares en campañas de desinformación. También se documentan 44 casos de IA generativa aplicada a fines políticos, con audios y videos alterados que afectan candidatos como Claudia Sheinbaum y Samuel García, quien tuvo que desmentir públicamente amenazas falsamente atribuidas a su voz (Rios Gutiérrez, 2024; Sprinforma, 2024).

Frente a este desafío, en México destaca el trabajo de laboratorios como Signa Lab ITESO y colectivos verificadores como Verificado 2018 y El Sabueso de Animal Político, que han analizado patrones de ataque, documentando la presencia de redes como la #RedAMLO y la Red Brolan. La #RedAMLO se organizaría en torno a productores de contenido ("maestros de ceremonias"), coros amplificadores, trolls y fans auténticos, impulsando tendencias y defendiendo la imagen presidencial de López Obrador en Twitter y otras plataformas (Vargas Pasaye, 2022; Infobae, 2020).

En 2025 el periódico *The New York Times* alertó sobre campañas por parte de medios propiedad del Kremlin ruso que intentan avivar sentimiento antiamericano en México, especialmente desde 2024 tras el bloqueo del canal RT (anteriormente *Russia Today*) en Europa y Estados Unidos (Abi-Habib, 2025), llegando a verse este canal en directo en el transporte público de México D.F. así como de otras ciudades mexicanas (Frias Deniz, 2025), otorgando también reconocimientos por parte del Club de Periodistas de México a periodistas vinculados a Rusia e Irán (Valencia, 2023).

1.7.3. Venezuela

A través del Ministerio de Comunicación e Información (MIPPCI) y del Sistema Bolivariano de Comunicación e Información (SIBCI), el Estado coordina diariamente campañas en redes sociales con etiquetas oficiales amplificadas por miles de bots y usuarios pagados mediante el Sistema Patria (Cañizález, 2021). Este mecanismo convierte la difusión de propaganda en un modo de subsistencia para miles de ciudadanos, mientras Twitter y otras plataformas han identificado y eliminado miles de cuentas vinculadas a estas operaciones. Solo en 2021 se calcula que generó 186,7 millones de mensajes en Twitter (Cazadores de Fake News, 2021; Bloomberg Línea, 2021). En mayo de 2022 los "Tuiteros de la Patria" llegaron a organizar una protesta en redes con las etiquetas #RespetoParaLosTuiteros y #TuiterosPatria porque el gobierno retrasó el pago de sus bonos (Correo del Caroní, 2022).

El sistema venezolano combina automatización tecnológica e incentivos en operaciones incluyen el uso de bots, cuentas "*cyborgs*", campañas de acoso a

periodistas y narrativas distractoras para moldear la conversación pública. Su estructura, según documentos filtrados expuestos en un informe del grupo Iberifier seguían una estructura militar, donde cada persona podía gestionar 23 cuentas, formando escuadras de 10 personas, compañías de 50, batallón de 100 y brigada de 500 (hasta 11.500 cuentas). Este mismo informe destaca campañas que las campañas de Venezuela se dirigieron a Estados Unidos y Europa (probablemente España) en Twitter, las de Bolivia a México y Venezuela, y las de Ecuador a Argentina, Chile o El Salvador (Badillo, & Arteaga, 2024). Los ataques de desinformación entre distintos países en la zona (Colombia, Perú, México, etc.) tiene abundantes ejemplos (Globalamericans, 2021), así como en el uso de IA y *deepfakes* (Singer, 2023).

Venezuela recibe también ataques contra su régimen, como por ejemplo una red de cuentas de ultraderecha desde España, poco antes de 2017 (Applebaum, 2020). En paralelo, el país formaría parte de una red transnacional de desinformación junto a otros países operando etiquetas y campañas coordinadas, pudiendo observarse como el 32% de redifusión de los canales rusos RT y Sputnik se producen desde este país según un informe oficial francés, siendo el segundo distribuidor tras España (Vilmer, & et al., 2018). Aunque existen actores opositores en el país que también disponen de producción de contenidos falsos en líneas opuestas al régimen, su capacidad es bastante menor frente a la maquinaria principal (Universidad de Palermo, 2023).

Frente a ello, han surgido iniciativas de verificación y coaliciones periodísticas (como el Observatorio Venezolano de Fake News o C-Informa) que intentan contrarrestar el fenómeno. El estudio "El orden global de la desinformación" de la Universidad de Oxford (2019) ubicó a Venezuela entre los ocho países (junto con China, India, Irán, Pakistán, Rusia y Arabia Saudí) que utilizan propaganda computacional y tropas cibernéticas para operaciones de injerencia extranjera y manipulación para audiencias globales (Bradshaw, & Howard, 2019).

1.7.4. Francia

Francia ha sido un objetivo persistente de las operaciones de influencia rusas desde la Guerra Fría, cuando Moscú la consideraba el "eslabón débil" de la OTAN. Durante décadas, los servicios soviéticos (KGB y GRU) cultivaron redes de desinformación aprovechando el peso del comunismo francés y el antiamericanismo de ciertos sectores intelectuales (Quénel, 2023). Con la llegada de la era digital, esa estrategia se reactivó en 2015 con el ataque cibernético de falsa bandera al canal de televisión *TV5 Monde*, que marcó un salto cualitativo, al igual

que las primeras campañas de manipulación ligadas al *Brexit* y a la imagen de Emmanuel Macron antes de su llegada al poder (Applebaum, 2020).

Las elecciones de 2017 fueron el primer gran choque entre Francia y la maquinaria moderna de desinformación rusa. Operaciones como las *Macron Leaks* y falsos artículos de medios tradicionales buscaban erosionar su credibilidad, mientras Sputnik y RT amplificaban rumores sobre vínculos con Estados Unidos o Arabia Saudita con documentos difundidos en la red social 4chan (Jeangène Vilmer, & et al., 2018). La respuesta del gobierno francés provocó la exclusión de esos medios del Elíseo y la creación, pocos años después, de mecanismos institucionales como Viginum en 2021, organismo encargado de detectar injerencias extranjeras, y la doctrina militar L2I, destinada a contrarrestar la guerra informativa.

Desde la invasión rusa de Ucrania, Francia ha enfrentado una escalada de campañas sofisticadas, como suplantaciones digitales de grandes medios e instituciones (*Doppelgänger*), falsificación de portales gubernamentales, y maniobras híbridas como el marcado de Estrellas de David en París para agravar tensiones internas. En paralelo, redes como *Portal Kombat* o *Storm-1679* intensificaron sus operaciones durante los Juegos Olímpicos de París 2024 y las elecciones europeas con la operación *Matriochka*, utilizando incluso inteligencia artificial para generar falsos documentales o comunicados (Viginum, 2024).

El Sahel ha sido testigo de una guerra de información estratégica en la que Rusia, a través de Prigozhin y el Grupo Wagner, ha desplegado campañas dirigidas a desacreditar a Francia y desplazar su influencia. Aprovechando el resentimiento colonial y el descontento social, las operaciones rusas han propagado narrativas sobre el neocolonialismo francés, promovido conspiraciones como la supuesta creación francesa de grupos yihadistas, y estimulado protestas antifrancesas mediante pagos directos y propaganda visual amplificada en medios rusos. Estos esfuerzos han impactado físicamente la región, como sucedió en 2021 con el bloqueo de un convoy francés en Burkina Faso, y han incluido operaciones de falsa bandera para acusar a las tropas galas de crímenes, desenmascaradas posteriormente por parte de la inteligencia francesa (Quénel, 2023). En respuesta se lanzarían campañas digitales que imitaban las tácticas rusas y protagonizó el primer "troll contra troll" documentado en la región, pero esa estrategia igualó las tácticas propagandísticas, dañando su propia legitimidad y consolidando la ventaja rusa en la narrativa, además de acabar posteriormente con la retirada francesa militarmente de la zona (Iriarte, 2025).

1.7.5. Alemania

Entre 2015 y 2017 Alemania enfrentó importantes operaciones de desinformación que aprovecharon la crisis migratoria para alimentar tensiones internas. Un caso emblemático fue el de Lisa, una adolescente que afirmó haber sido atacada por hombres de origen "meridional", un incidente que fue rápidamente explotado por medios y redes rusas como *Sputnik* y *RT Deutsch*, así como por movimientos neonazis y antiinmigrantes como PEGIDA (Colon, 2025). Aunque la policía desmintió el caso, la campaña propagandística continuó, exacerbando el sentimiento anti-migrante no solo en el país sino en toda Europa. A esta manipulación se sumó a la campaña electoral de 2017 del partido de extrema derecha Alternativa para Alemania (AfD), que centró su discurso en la inmigración y recibió respaldo tácito del Kremlin, según acusaciones oficiales, influyendo en los resultados electorales mediante la difusión de mensajes negativos contra Angela Merkel.

Con la guerra en Ucrania, las operaciones de desinformación se volvieron aún más sofisticadas, destacando la campaña *Doppelgänger* que suplantó a medios tradicionales alemanes de prestigio como *Spiegel* y *Die Welt* (de Pedro, & et al., 2023). Esta estrategia buscó principalmente erosionar el apoyo a las sanciones contra Rusia y a la recepción de refugiados, utilizando temas sensibles como la crisis económica y migratoria para dividir a la sociedad alemana. Investigaciones revelaron que hasta un 80% de los comentarios en redes sociales en Alemania y Austria tenían una inclinación prorrusa, y la suplantación digital se extendió incluso a organismos oficiales como el Ministerio del Interior, como parte de un esfuerzo sistemático de manipulación y desestabilización (Jeangène Vilmer, & et al., 2018).

En el contexto electoral de 2024, la campaña *Doppelgänger* continuó con páginas falsas creadas para influir en las elecciones europeas, apoyadas por anuncios pro-Kremlin aprobados por plataformas como Meta, dirigidas a países clave de Europa occidental. El Gobierno llegó a calificar estas acciones como una guerra informativa destinada a crear fracturas políticas y sociales profundas (European Union External Action, 2024). Un informe sobre injerencias extranjeras en las elecciones federales de Alemania de 2025 revela una campaña sostenida, principalmente de origen ruso, que explotó divisiones sociales y temas como migración, economía y seguridad usando narrativas manipuladoras, contenido generado por inteligencia artificial, suplantación de instituciones y redes de bots. Estas tácticas fueron adoptadas incluso por actores domésticos como el

AfD, mientras las brechas regulatorias y de aplicación en plataformas permitieron la persistencia y amplificación de dichas operaciones, complicando cada vez más diferenciar entre manipulación extranjera y doméstica (Institute for Strategic Dialogue, & et al., 2023).

1.7.6. Filipinas y Vietnam

Filipinas se ha consolidado como el epicentro global de la desinformación comercializada, donde la industria opera con una estructura piramidal profesionalizada. Sus estrategas de publicidad y relaciones públicas diseñan campañas que son ejecutadas por *influencers* y una vasta base de operadores digitales mal pagados, muchos de ellos jóvenes universitarios en situación precaria (0,45 dólares por publicación). Este ecosistema, revelado por investigaciones como la de *Rappler* y *Architects of Networked Disinformation*, se nutre de modelos flexibles que permiten a políticos, empresas y clientes internacionales contratar paquetes de viralización con métricas garantizadas de alcance y *engagement* (Ong, & Cabañes, 2019; Ong, Tapsell, & Curato, 2019). Ejemplos como el de *Twinmark Media* muestran que estas consultoras pueden movilizar decenas de miles de dólares mensuales, proporcionando influencia digital y servicios a escala para campañas políticas, comerciales y de manipulación social, extendiendo el modelo filipino a otros países del sudeste asiático (Hapal, 2024).

Mientras tanto, Vietnam ha optado por una estrategia estatal y militarizada a través de Force 47, conformada por al menos 10.000 efectivos. Este grupo actúa principalmente en Facebook y otras plataformas digitales con el objetivo explícito de combatir puntos de vista considerados "erróneos" por el régimen comunista, promoviendo contenido oficialista y acallando voces disidentes. Las operaciones incluyen reportes masivos coordinados para eliminar publicaciones, acoso digital y doxeo (publicar información personal) de activistas, además de la manipulación sistémica de algoritmos y cuentas anónimas para influir en la narrativa pública. El uso de presión sobre empresas como Facebook ha resultado en mayor censura de contenido político, y la implementación de leyes de ciberseguridad y acceso a datos personales ha fortalecido el control digital del Estado.

La sofisticación y escala de las operaciones en ambos países han generado preocupación internacional, pues Filipinas exporta técnicas y servicios a campañas digitales de otras naciones, mientras Vietnam perfecciona el aparato represivo en el entorno *online*. Desde la persecución judicial a activistas (como Le Van Dung, condenado por difundir contenidos críticos) hasta el funcionamiento de cientos de granjas de trolls en Manila capaces de escalar campañas

virales transnacionales. Ambos modelos ilustran la convergencia entre desinformación comercial y control político autoritario, convirtiéndose en laboratorios de manipulación informativa y represión digital.

1.7.7. África

El Sahel africano se ha convertido en el epicentro estratégico de las campañas de desinformación en el continente, siendo Rusia el principal actor externo detrás de este fenómeno. Desde 2018 a 2024, se han identificado al menos 19 campañas rusas dirigidas específicamente a Malí, Burkina Faso y Níger, coincidiendo con golpes de estado militares y el auge del sentimiento antifrancés y antioccidental, proceso impulsado a través de una red multicanal que conecta medios estatales como RT y *Sputnik*, mercenarios del Grupo Wagner y nuevas plataformas como *African Initiative* o la red de sitios web *Portal Kombat* (Gutiérez, 2024; Secrétariat Général de la Défense et de la Sécutité Nationale-SGDSN, 2025; EUvsDisinfo, 2025). Esta reconfiguración, acelerada tras la muerte de Prigozhin en 2023, refuerza el control ruso sobre la propaganda y la gestión de narrativas, trasladando operaciones del grupo Wagner al llamado *Africa Corps* (Pérez Triana, 2025).

Las estrategias rusas en la región emplean la táctica del "blanqueo de contenidos", que consiste en diseminar información manipulada por canales locales aparentemente legítimos, incrementando su credibilidad ante las audiencias africanas. Las narrativas aprovechan el discurso anticolonial y panafricanista, presentándose como aliado frente al pasado occidental, y amplifican la percepción de decadencia moral de Occidente y el fracaso de sus intervenciones en la región (Bencherif, & Carignan, 2023). El alcance de estas campañas se potencia por el alto consumo de redes sociales y la juventud promedio africana, convirtiendo a la población más joven en objetivo para la manipulación informativa que legitima a juntas militares y reconfigura las alianzas geopolíticas (Fuente Cobo, 2025; TruthAfrica, 2025). Los canales de emisión de noticias rusos se convirtieron en 2018 entre los más vistos en países como Libia, Túnez, Yemen o Egipto, mediante acuerdos con medios locales (Disinfo África, 2023).

Aunque Rusia lidera el panorama de la propaganda algorítmica en África, países como China, Turquía, Emiratos Árabes Unidos y Arabia Saudí también despliegan las suyas propias, muchas veces mediante inversiones mediáticas y conexiones culturales o religiosas (Blanco, & Abril, 2023) en casos como la guerra en Sudán (Jones, 2023). Además, actores locales (partidos políticos, líderes e incluso grupos yihadistas como JNIM y EI Sahel) utilizan tácticas de comunica-

ción para influir en sus comunidades y consolidar simpatías. Las debilidades estructurales del ecosistema informativo africano, el uso generalizado de WhatsApp y radios locales, y la escasa moderación de contenido en lenguas africanas por parte de plataformas como Meta facilitan la implantación y el éxito de estas campañas, potenciados por frecuentes apagones digitales impuestos por gobiernos locales durante momentos críticos como elecciones y protestas.

Ejemplos como la intervención de CitizenGo, organización ligada a la ultraconservadora y ultracatólica Hazteoir de España, en las elecciones de Kenia en 2022 ilustran las muchas injerencias en la zona (Shiundu, & Jiménez, 2022). Por otro lado, Meta ha eliminado cuentas y redes sociales no solo a los actores anteriores, sino también asociadas a Francia que participaban en campañas en el Sahel, difundiendo narrativas favorables a Francia y críticas a Rusia y a los gobiernos militares africanos (Meta, 2020). Francia terminó retirándose de la zona en 2022 de Mali y en 2023 de Burkina Faso y Níger (Holyoke, 2025).

Pero no solo se recibe desinformación en África, sino que también es uno de los mayores centros productores. Por ejemplo, las *Yahoo Yahoo Schools* de Nigeria son centros clandestinos donde jóvenes, incluso menores de edad, reciben entrenamiento intensivo en técnicas de fraude en línea, ingeniería social, manipulación psicológica, uso de *scripts* y *deepfakes*, con el objetivo central de estafas internacionales como la extorsión sexual y los fraudes románticos (Lazarus, 2018). Estas organizaciones operan tanto en espacios físicos como en foros digitales y han extendido su influencia a nivel global, afectando a víctimas en América, Europa y Asia, mientras que plataformas como Meta han cerrado decenas de miles de cuentas vinculadas a estas redes. Aunque su actividad principal sigue siendo económica, disponen de capacidades técnicas para participar en campañas de desinformación al servicio de distintos clientes (SimplVest, 2025). Numerosas cuentas en noviembre de 2025 fueron expuestas en X al publicarse el lugar desde donde emitían, apareciendo así multitud de cuentas de apoyo MAGA en Estados Unidos con sede en Nigeria o Kenia, junto a otros países del sudeste asiático como India, Bangladesh o Tailandia, o el este de Europa (Orr Bueno, 2025).

1.7.8. China

La estrategia de propaganda de la República Popular China (RPC) tiene profundas raíces históricas, aceleradas desde la elevación de su oficina de Propaganda Exterior a nivel ministerial en 1991. En los últimos años la información se consolidó como prioridad estatal, con una inversión de alrededor de 6.600 millones de dólares desde 2009 para fortalecer su influencia mediática global (South

China Morning Post, 2009). China busca controlar y moldear el flujo informativo tanto dentro como fuera de sus fronteras como apoyo a sus objetivos estratégicos, y uno de los instrumentos más antiguos de esta misión es la "Armada de los Cincuenta Céntimos" (*Wumao*), una fábrica de trolls activa desde 2004 que publica millones de mensajes para desviar la atención de temas sensibles al régimen y modelar la conversación digital, llegando a ocupar millones de individuos y evolucionando hacia nuevos fenómenos como los *ziganwu*, blogueros hipernacionalistas (Colon, 2025; Iriarte, 2025).

En el escenario internacional, apoyado en el concepto de *sharp power* (mientras que el *soft power* busca influir mediante cultura o vía diplomática, el *sharp power* distorsiona la información) para penetrar y manipular los ecosistemas digitales rivales mediante tácticas industriales que combinan la acción de ejércitos humanos con sofisticados recursos algorítmicos y de inteligencia artificial, como en las elecciones de Taiwán en 2020 (Huang, 2025). Redes como *Spamouflage*, operan desde al menos 2014 distribuyendo contenidos y desinformación sobre temas críticos (protestas en Hong Kong, Covid-19, etc.) en plataformas globales (Colon, 2025; Graphika Atlas, 2025). A esta maquinaria se suma el papel de TikTok, controlada por ByteDance, que no solo recopila datos, sino que actúa como canal de propaganda directa y de censura selectiva, suprimiendo contenidos sobre minorías o conflictos y promoviendo narrativas afines al régimen, como se evidenció durante la guerra de Ucrania y en el silenciamiento de información sobre Xinjiang (Iriarte, 2025).

En el plano tecnológico, China lidera ya el despliegue de soluciones basadas en inteligencia artificial para la generación automatizada de contenidos y campañas en procesos electorales de otros países. El uso de *deepfakes* para desacreditar a opositores en elecciones clave, como el caso de Taiwán en 2024, marca el salto a una "guerra cognitiva" donde la creación de falsificaciones de audio y video se convierte en una herramienta político-diplomática central (European External Action Service, 2024).

En el informe francés de finales de 2025 del *Institute de Recherche Stratégique de l'École Militaire*, titulado *Baybridge, anatomy of a chineses information influence ecosystem* (Tadaweb, & Charon, 2025), se exponen una red de operaciones de las empresas *Shenzhen Haimai Yunxiang Media* y *Shanghai Haixun Technology* con, por un lado fines comerciales pero por otro la presencia de más de 100 sitios webs de noticias falsas en una docena de idiomas, dirigidos hacia 30 países. Sus campañas se enmarcan principalmente dentro de la expresión "energía positiva" (*zheng nengliang*) a la hora de transmitir mensajes, en un tono optimista uniforme de difusión de la posición china, pero de crítica hacia países occidentales.

Se identifica así la operación *HaiEnergy* principalmente hacia audiencias estadounidenses en 2022, la operación *Paperwall* en 2024, o la *SPR Network* hacia Asia Sudoriental, apreciando en los últimos años una fuerte inclinación a la narrativa pro-Kremlin y su "operación especial" en Ucrania en países como Francia. En los análisis se detectaron 241 campañas de aspecto político hacia Asia (28%), 193 hacia Europa (22%, con 21 hacia Reino Unido, 20 a España, 18 a Italia, 17 a Portugal, 15 a Francia y otras 15 a Alemania, entre otros), 92 hacia Latinoamérica (11%) y 59 hacia Estados Unidos (7%). A pesar de su gran extensión, está valorada como bastante ineficaz por su baja calidad del contenido, malas traducciones automáticas, elementos irrelevantes, prioridad de cantidad sobre calidad, así como narrativas contradictorias y descoordinadas sin entender el contexto cultural de la sociedad objetivo.

En el informe de agosto de 2025 de la empresa Graphika se presenta la operación "Falsos Amigos" (Fulde-Hardy, 2023) (literalmente en español, ya que los nombres informáticos empleados se llamaban "amigos"), que fue una red coordinada que comprendió 11 dominios y 16 cuentas o páginas en múltiples redes sociales como Facebook, X, Instagram, Mastodon y Threads, y cuyo objetivo principal era lavar y diseminar informes del medio de comunicación estatal chino CGTN. Los activos de la red habrían empleado herramientas de inteligencia artificial para traducir y resumir los artículos originales en inglés a varios idiomas, incluyendo español, francés y vietnamita, en un intento de ocultar la verdadera fuente del contenido. Se generó contenido con un estilo narrativo floreado y rimbombante, así como un uso grande de *emojis*, logotipos y perfiles diseñados para audiencias jóvenes en África, América, Asia y Europa que sugiere uso de IA. El mensaje difundido era principalmente pro-chino y anti-occidente, con una coordinación técnica extrema realizada en dominios registrados a través de *Alibaba Cloud Computing Ltd.*, diez de los cuales situaban su ubicación en Beijing. A pesar de esta sofisticación y del uso de anuncios pagados en Meta, la red no logró obtener éxito en redes sociales, recibiendo casi nulo *engagement* y consiguiendo unas pocas docenas de seguidores, aunque algunos de sus mensajes sí aparecieron en los resultados de búsqueda superiores de X para ciertos temas.

Aunque últimamente se detecten sinergias entre campañas pro-Kremlin y chinas, históricamente siempre existieron diferencias, ya que mientras que la guerra psicológica rusa podría resumirse en "sumisión a través de la confusión" y donde la reputación no importa en absoluto, China en cambio ha buscado siempre mantener una imagen de poder "moral y fiable". Pero desde la firma de un acuerdo de cooperación en información en 2021, la red *Spamouflage* está mostrando cada vez más técnicas desarrolladas en Rusia. Sus últimas narrativas

empleadas incluso se les denominan MAGAflage, ya que imitan el estilo y el lenguaje de la extrema derecha estadounidense (Sobchuk, 2025).

1.7.9. Israel

Israel es identificado como una de las tres grandes potencias pioneras en la manipulación de información por internet, en la que se emplea la *Hasbará* ("explicación" en hebreo), especialmente tras la Operación Plomo Fundido de 2009, creando una red de personas que postean respuestas proisraelíes. Su desarrollo, a imitación del nombre de la cúpula antimisiles del país, fue la creación de la herramienta *Words of Iron* (palabras de hierro) para operar masivamente en redes sociales (Iriarte, 2025).

Tras el ataque de Hamas a colonos israelíes en 2023, se produjeron multitud de campañas en redes sociales a favor de este grupo, así como campañas rusas que decían que las armas para matar judíos habían sido suministradas por el presidente ucraniano Zelensky o que fue una operación de falsa bandera, que fueron respondidas en Israel con publicidad *online* en 26 países con 7,1 millones de dólares, especialmente Francia (3,8 millones) que había sido objetivo de la campaña antisemita RRN por parte de cuentas del ecosistema ruso (Colon, 2025; Vilmer, & et al., 2018).

Las operaciones de influencia de Israel se han transformado en una maquinaria sofisticada y multimillonaria, impulsada por el Ministerio de Asuntos Exteriores, con un enfoque primario en la opinión pública de Estados Unidos. Recientes documentos exponen cerca de 150 millones de dólares aprobados a finales de 2024 con ese fin (Jones, 2025a). Esta vasta campaña se gestionó a través del gigante publicitario *Havas Media Group* y una red de firmas de relaciones públicas estadounidenses con buenas conexiones, contratando a empresas alineadas tanto con el Partido Republicano (como *Clock Tower X LLC*) como con el Partido Demócrata (como *SKDKnickerbocker LLC*), con el objetivo de inyectar narrativas pro-Israel en los niveles políticos más altos de EE.UU. (Jones, 2025b). Una estrategia incluyó un contrato de 6 millones de dólares dedicado a una campaña impulsada por IA para conquistar a la Generación Z, destinando el 80% del contenido a plataformas como TikTok, Instagram y YouTube.

Los estudios del académico Owen Jones de 2025 revelaron la intención de contratar a una firma para implementar un "programa basado en bots" en múltiples plataformas sociales, incluidas Instagram, TikTok y YouTube, cuyo propósito era "inundar la zona" con mensajes pro-Israel. Además, la campaña incluía un componente hiper-segmentado dirigido a las comunidades cristianas

estadounidenses con una inversión de 3,2 millones, planificaba las campañas en propiedades de iglesias y *colleges* cristianos durante los horarios de culto. Esta maquinaria de influencia también recurrió al pago directo a *influencers* de redes sociales, con tarifas de hasta 7.372$ por publicación, y buscó activamente manipular los propios modelos de lenguaje de las IA, intentando así proteger la reputación de Israel (Jones, 2025a). Hacia Europa se detectaron campañas de 2 millones de dólares para negar la hambruna en Gaza en 2025 (Wesolowski, 2023), así como los ataques entre Irán e Israel de junio de 2025 se llenaron de campañas en ambas direcciones (Abdel Aziz, 2023).

En contraste con sus operaciones ofensivas, Israel fue objeto de una oleada de desinformación masiva, especialmente durante la Guerra entre Israel e Irán de junio de 2025, denominada "Falsificaciones Balísticas" (*Ballistic Fakes*) (Yasur, Ram, & Ring, 2025). La motivación principal, que potencialmente sirvió a intereses iraníes, fue transmitir poder y exagerar la superioridad militar y la destrucción infligida al enemigo, centrándose en temas de daño físico a edificios, explosiones y despliegues de fuerza militar. La mayoría de este contenido era visual, predominantemente vídeos (84%), y el método de manipulación más común fue la decontextualización de eventos auténticos (88%), como el uso de clips de ataques de misiles pasados o vídeos de China, Estados Unidos o Rusia para simular la destrucción en curso. El rol de la IA en los ataques contra Israel fue significativo, constituyendo el 20% del contenido falso verificado durante el conflicto, y se utilizó para generar imágenes y vídeos falsos que simulaban daños físicos y la ruina de infraestructuras en Israel, incluyendo el colapso de edificios o el Aeropuerto de Tel Aviv en ruinas. Además, aunque Israel es uno de los líderes mundiales en el uso de redes sociales, su ecosistema de verificación de hechos es limitado y subdesarrollado, enfrentando presiones financieras y políticas (Ram, Wiener, & Ring, 2025).

Team Jorge, que saltó a la luz pública por investigaciones periodísticas en 2023, es un grupo mercenario digital israelí integrado por exmilitares y expertos en inteligencia, cuyo negocio principal consiste en manipular procesos electorales y campañas políticas a escala internacional mediante *hackeos*, chantaje y grandes campañas de desinformación (Bakir, & Briant, 2024). Actúan bajo encargo tanto de estados como de actores privados, pero rehúsan operaciones en elecciones israelíes, estadounidenses o que puedan disgustar al gobierno ruso. Entre los casos documentados se encuentra su intervención en Nigeria (2015) junto a Cambridge Analytica, usando correos manipulados y chantajes, y su participación en las elecciones de Kenia (2018), donde emplearon multitud de perfiles falsos y difusión tipo *astroturfing* (método de difusión que se explicará más

adelante), amplificando historias menores hasta convertirlas en virales y ejecutando operaciones conjuntas con periodistas para forzar despidos y alterar el debate público en al menos 30 países (Emergui, 2023; Kirchgaessner, & et al., 2023). Además de manipular a la opinión pública, Team Jorge ofrece igualmente servicios de borrado reputacional y eliminación de informaciones perjudiciales, muchas veces en coordinación con otras firmas especializadas. Estas empresas están implicadas en cientos de operaciones para eliminar noticias negativas sobre clientes investigados o condenados por corrupción, narcotráfico y delitos financieros, utilizando tácticas ilícitas como la creación de webs falsas, denuncias por derechos de autor simuladas y manipulación de buscadores.

Otras empresas israelíes de las que se han conocido operaciones han sido *Psy-Group* (especializada en operaciones de inteligencia privada, campañas psicológicas y manipulación digital, investigada por su oferta de servicios a la campaña de Trump y sus relaciones con Cambridge Analytica; cerrada tras las revelaciones periodísticas de 2018, aunque parte de su personal migró a otras consultoras tecnológicas israelíes), *Black Cube* (consultora de inteligencia privada compuesta por antiguos miembros de agencias israelíes, conocida por campañas influyentes y trabajos oscuros para gobiernos y grandes corporaciones, citada en investigaciones sobre manipulación de información en procesos de alto perfil globales, incluidas campañas de desprestigio y recopilación de datos sensibles), *NSO Group* (aunque principalmente famosa por su programa Pegasus para espiar teléfonos móviles, ha sido relacionada en estrategias de vigilancia, recopilación masiva de datos y potencial influencia en campañas informativas), *Wikistrat* (que llegó a realizar simulaciones de operaciones en Yemen, cerró en 2017), o *Inspiration* (que llegó a ofrecerse en la primera campaña de Trump en Estados Unidos para operar sobre votantes de estados clave) (Siegelman, 2018).

1.7.10. Brasil

Brasil se ha consolidado como uno de los casos más críticos de desinformación en el Sur global, donde la extrema derecha ha explotado masivamente plataformas como WhatsApp y Facebook para atacar la confianza en las instituciones democráticas (Recuero, Guazina, & Araújo, 2025). En las presidenciales de 2018, la campaña de Jair Bolsonaro se apoyó en redes de "máquinas de noticias falsas" que enviaban hasta mil mensajes diarios por grupo. Esa dinámica continuó en 2022, cuando Bolsonaro y su entorno sembraron dudas sobre la fiabilidad de las urnas electrónicas desde sus perfiles personales, mucho más seguidos que las

cuentas institucionales, trasladando rumores y teorías conspirativas al centro del debate público (Applebaum, 2020).

La creación de la llamada "Oficina del Odio" profesionalizó la producción y coordinación de ataques digitales contra opositores, periodistas y jueces, contribuyendo al clima que desembocó en el asalto a las instituciones en Brasilia en enero de 2023 (Santini, & Salles, 2025). En paralelo, se ha desarrollado una infraestructura propagandística estable, ejemplificada por la productora *Brasil Paralelo*, que desde 2016 difunde documentales revisionistas como "1964: O Brasil entre Armas e Livros", minimizando la dictadura militar y ofreciendo una "historia alternativa" con fuerte sesgo. Dichas campañas combinan contenido verídico junto con narrativas distorsionadas y campañas agresivas en redes, usando una lógica de "manguera de falsedades" para erosionar la credibilidad de medios tradicionales y crear una alfabetización mediática de extrema derecha que enseña a desconfiar sistemáticamente de las élites. Al mismo tiempo, Brasil ha sido epicentro regional de desinformación sanitaria durante la pandemia de Covid-19, por donde circularon masivamente remedios falsos, mensajes negacionistas y ataques a las vacunas, muchos amplificados por el propio Bolsonaro, que emitió más de mil declaraciones falsas sobre el virus según verificadores como *Aos Fatos* (Assis, 2023), mientras Brasil originaba el 9% de los bulos sobre Covid-19 recopilados por la red Latam Chequea (Badillo, & Arteaga, 2024).

Estas dinámicas se ven potenciadas por la altísima penetración de redes y mensajería, pues Brasil es uno de los principales países del mundo en usuarios de Facebook y cuenta con unos 120 millones de usuarios de WhatsApp, además de ser la principal región del mundo de noticias vistas a partir de las redes sociales, lo que convierte estas plataformas en canales privilegiados para la difusión de rumores y campañas coordinadas (Jeangène Vilmer, & et al., 2018).

1.8. ESPAÑA

La mayor parte de las informaciones y referencias que salen en los medios de comunicación (las pocas que salen), generalmente hacen mención a situaciones que suceden fuera de las fronteras españolas. España en el informe del grupo de Cambridge sobre Propaganda Computacional en 2020 no aparece como un país donde se perciban grandes industrias dedicadas al efecto, de hecho se le considera de nivel de emisión bajo. Pero eso no significa que ni emita ni que tampoco que lo sufra, siendo además el tercer país del mundo que más investigaciones académicas tiene al respecto (Rodríguez Fernández, & et al. 2023).

Uno de los casos más sorprendentes, y poco conocidos, es el que sitúa uno de los posibles orígenes de esta industria mundial justamente en España. Así el diario El Periódico publicó en agosto de 2025 (Fernández, & Calleja Flórez, 2025) un reportaje relacionado con el *software* "Social Baton", que tenía un costo aproximado de 100.000 euros. Este *software* creaba conversaciones como si fueran auténticas, analizaba redes sociales, identificaba y clasificaba a los más influyentes, segmentaba audiencias con el fin de poder realizar campañas de desmovilización, desprestigio, contraataque e influencia social. Fue distribuido por el empresario Javier Pérez Dolset en 2015 a clientes selectos, según este medio, para investigar y controlar a la opinión pública, entrando en contacto con el Partido Popular y el PSOE, en un momento crucial en el que Mariano Rajoy se jugaba la reelección como presidente del Gobierno. Se reconoció una reunión con la Secretaria de Estado de Comunicación de la época en la sede del PP en la calle Génova, mientras que Susana Díaz (aspecto que ella niega que empleara) competía en las primarias contra Pedro Sánchez en el PSOE.

Social Baton se publicitaba como *made in Rusia* a través de la empresa Zed, propiedad de Pérez Dolset, que contaba con uno de sus centros de desarrollo en Rusia desde 2006 especializado principalmente en música y videojuegos (empresa que creó Teleline, que posteriormente sería Terra, así como Pyro Studios con su famoso videojuego Commandos). Tras la campaña del partido Convergència i Unió (CiU) en 2010 que utilizó intensamente la digitalización, y los eventos del movimiento 15M, los partidos comenzarían a emplear programas como Social Baton, que incluía en su promoción un dosier publicitario con una recomendación del director de campaña *online* del PP. En 2011, se habría vendido una versión del programa al Secretario de Estado de Estados Unidos en 2025 Marco Rubio, cuando era candidato por Florida.

Esta empresa aprendería de la experiencia de Cambridge Analytica y ampliaría su catálogo de servicios. Pérez Dolset estableció contactos en Rusia con varios interesados en controlar las redes sociales, entre ellos la *Internet Research Agency* (IRA) del grupo Wagner, liderada por Yevgeny Prigozhin, así como altos funcionarios del gabinete de Putin. En marzo de 2014, las autoridades rusas retiran los visados a Pérez Dolset y a sus ejecutivos españoles, lo que derivó en la pérdida de sus oficinas en Moscú y San Petersburgo. Mientras era investigado por la Audiencia Nacional española, el dueño de la compañía denunció que le habrían robado las plataformas *Social Baton* y *Glass Eye*. Esta última se utilizaba para rastrear *routers* y usuarios en la vida civil, y su tecnología era muy similar a la empleada en el sistema de inteligencia ruso SORM-3.

Otras informaciones han aparecido a lo largo del tiempo:

- Tras el derribo del vuelo MH17 sobre Ucrania en julio de 2014, donde murieron 298 personas, Rusia puso en marcha una sofisticada campaña de desinformación para confundir sobre la autoría del ataque. Los medios rusos difundieron primero la noticia de que se habría tratado de un avión militar ucraniano y, al cambiarse la versión al descubrirse que era un vuelo civil, introdujeron rápidamente a "Carlos Spainbuca", un supuesto controlador aéreo español que afirmaba la presencia de cazas ucranianos junto al Boeing malasio y que supuestamente estaba amenazado por su opinión crítica a las protestas del Maidán en Ucrania. Este personaje, en realidad un ciudadano español financiado por Rusia, se convirtió en el eje de la narrativa alternativa impulsada por medios y redes sociales, llegando incluso a ser citado tiempo después por el propio presidente Putin en una entrevista con Oliver Stone, a pesar de que su identidad y falsedad ya habían sido desenmascaradas (de Pedro, & et al., 2023).
- En octubre de 2018 Twitter divulgó un listado de cuentas que atribuyen a operaciones de influencia por parte de gobiernos de distintos países del mundo, con 259 cuentas ligadas al de España y 130 al de Cataluña (Jones, 2025).
- En enero de 2019 Meta eliminó 783 páginas, grupos y cuentas con conexiones con Irán con actividad específica en diversos países del mundo, incluida España. Su principal contenido es la introducción de temas entre Israel y Palestina, Siria y Yemen, ocultando su procedencia y haciéndose pasar por personas y grupos locales de cada sitio (Meta, 2019a).
- En septiembre de 2019 Meta elimina, a partir de contactos con Twitter, 65 cuentas de Facebook y 35 de Instagram por "contenido y comportamiento no auténtico" procedentes del Partido Popular para hacerse pasar por otras personas (Meta, 2019b).
- En mayo de 2019 Meta elimina 16 cuentas de Facebook, 4 páginas y una cuenta de Instagram por contenido coordinado no auténtico procedente de Rusia con objetivo en varios países europeos, incluida España, en la que se introducen temas en torno a inmigración, religión y en contra de la OTAN (Meta, 2019c).
- En 2019 sale a la luz la empresa Cenyt, perteneciente al comisario José Manuel Villarejo, que ofrecía entre sus servicios "maniobras de intoxicación informativa" para "generar desconcierto y/o desconfianza". Se

habría realizado un trabajo ("Informe King") para una de las exposas del dictador de Guinea Ecuatorial Teodoro Obiang, para desacreditar en distintas plataformas a uno de los hijos que tenía con otra mujer, y asegurar que le sucediese su otro hijo (Iriarte, 2025).

- En septiembre de 2020 Meta elimina una página y 5 cuentas de Facebook originadas en Rusia que se hacían pasar por ser residentes en otros países introduciendo conspiraciones geopolíticas, desinformación sobre elecciones parlamentarias, pandemia por Covid-19, brutalidad policial, injusticia social y racial y comentarios en varios países, incluída España (Meta, 2020).
- En diciembre de 2020 Meta elimina 23 cuentas de Facebook, 25 páginas, 11 grupos y 19 cuentas en Instagram que desde la región ucraniana de Luhansk, ocupada por Rusia, se lanzaron campañas de desinformación hacia España ligadas a una organización pro-rusa en el sur y este de Ucrania (Meta, 2021).
- En 2020, durante la pandemia por covid-19 se distribuye mundialmente muchísima desinformación y campañas contra diferentes vacunas, situando el grupo de verificadores Latam Chequea a España como el origen de una cuarta parte de las noticias falsas en español sobre el virus, identificando que el 15% de las noticias falsas presentes en un país no procedían del mismo. También se detectó que las plataformas de internet primaban la identificación de bulos en inglés frente a otros idiomas en dicho periodo. Igualmente la pandemia fue empleada para campañas de hispanofobia en grupos independentistas (Badillo, & Arteaga, 2024).
- En el 2021 la empresa IE3 Ventures, perteneciente a NicesTream, una consultora española, realiza una campaña contra varios jugadores y la oposición a la directiva del Fútbol Club Barcelona conociéndose en la prensa como el "Barçagate". Otra empresa del mismo grupo, Iluminati Lab, se conoce realizó tareas de propaganda digital en Ecuador, Chile, Argentina o España en distintos procesos políticos (Planas Bou, 2021).
- En 2023 un empresario israelí (conocido bajo el seudónimo Team Jorge) se atribuyó en una entrevista del grupo de medios *Forbidden Stories*, campañas en redes sociales que relacionaban al independentismo catalán con el Estado Islámico (Gil, & Irujo, 2023).
- En 2024 la empresa Meta eliminó en España 161.000 contenidos de desinformación en Instagram y unos 80.000 en Facebook en las semanas

previas y posteriores a las elecciones europeas del 9 de junio. Se catalogaron entre el 7 de mayo al 23 de junio solo en España un millón y medio de contenidos generados con IA en Facebook y 73.000 en Instagram. Igualmente, en el primer semestre de 2024 Google eliminó por contenidos desinformativos en España 1.207 canales, Tiktok 21.430 vídeos. También Tiktok tuvo que borrar 1.314 vídeos por ir contra la integridad cívica en época de elecciones y 270 por manipulación de medios, información facilitada por la Unión Europea (Swissinfo, 2023).

- En 2024 el Grupo de Acción Exterior de la Unión Europea alerta de la Operación Falsa Fachada, donde al menos 23 sitios web con contenido que posteriormente instrumentalizaba el ecosistema de propaganda y de desinformación pro-Kremlin, con uso de IA, hacia varios países, incluyendo España (European External Action Service, 2024).
- En 2024 sale a luz pública la existencia de un informe por parte del Ejército en el que se habla de un ecosistema de desinformación ruso con 179 focos emisores, para por un lado desinformar a la opinión pública española, pero también hacia la población rusa asentada en el país (Fernández, 2024).
- En 2025 el Informe Anual de Seguridad Nacional, correspondiente al año anterior, especifica que durante la DANA de Valencia en 2024 "el ecosistema de propaganda y desinformación pro-Kremlin, con carácter oportunista, amplificó y adoptó narrativas desinformativas preexistentes para su beneficio. Los actores pro-Kremlin se focalizaron en promocionar la desconfianza ciudadana en las instituciones públicas, deslegitimar el apoyo a Ucrania so pretexto de la necesidad real de ayuda a las zonas afectadas por la DANA y en proyectar una imagen de país sumido en el caos.", hechos que igualmente identificó en las elecciones al Parlamento Europeo (DSN - Presidencia del Gobierno, 2025).
- En enero de 2025 la empresa de análisis e inteligencia de EE.UU. denominada *Graphika*, atribuye operaciones encubiertas chinas en las que, haciéndose pasar por una ONG española llamada *Safeguard Defenders*, se solicita el derrocamiento del gobierno español por las inundaciones de la DANA de 2024 (The Graphika Team, 2025).

Como se puede comprobar Meta (propietaria de Facebook, Instagram o Whatsapp, por ejemplo) ha venido señalando sus operaciones contra la desinformación a lo largo de un periodo, del que no se puede encontrar ya ningún

tipo de actualización desde 2022. Justamente la Comisión Europea señaló en 2024 que no cumplia con su obligación por evitar la difusión de publicidad engañosa y campañas de desinformación, además de eliminar una herramienta de información pública que ayudaba en la lucha contra desinformación como era *CrowdTangle* (RTVE, 2024).

Según el explorador de industria de influencia, realizado por el *The Influence Industry Project* (The Influence Explorer, 2025), se pueden encontrar la participación de empresas que habrían efectuado campañas de influencia sobre la opinión pública en España. Así exponen a "Ridder/Baden", que trabaja para partidos políticos y empresas según su propia publicidad, "El equipo de campaña" que con sede en México tiene al frente a un consultor español, y en su web presumen de ganar el 93% de las campañas políticas (El equipo de campaña, n.d.), "*The Messina Group*" en el que incluye en su web a Mariano Rajoy como líder mundial cliente de la compañía junto a otros presidentes de otros países mediante campañas basadas en los datos, o la empresa "Liegey Muller Pons", que cambió su nombre a "eXplain" y que habría ejercido campañas de inteligencia algorítmica para el PSOE en 2015. De todas formas, este proyecto no coloca ninguna empresa conocida que ejecutara en España campañas con métodos y formas ilegales, frente a otros lugares del mundo.

De todas maneras otras empresas españolas, como "Eliminalia", han sido citadas por realizar campañas de influencia en otros países del mundo, especialmente Latinoamérica, por parte de entidades como el grupo *Computational Propaganda Research Project* de la Universidad de Oxford en su informe de 2020 (Bradshaw, Bailey, & Howard, 2020).

1.8.1. El proceso independentista en Cataluña

La intervención pro-Kremlin en el *procès* catalán comenzó a manifestarse en 2014, cuando medios como Sputnik empezaron a difundir desinformación sobre el movimiento independentista catalán, coincidiendo con el referéndum convocado ese año. Ese mismo período vio la creación de redes digitales que luego se centrarían en la polarización política de España, así como contactos directos entre políticos catalanes y representantes rusos en congresos en Moscú. En 2016, se documentó una campaña coordinada de desinformación por parte de RT y Sputnik, apoyada por bots y cuentas falsas, y circuló el bulo de que una Cataluña independiente reconocería a Crimea como rusa, amplificado por otros medios como Hispan TV (Alandete, 2022). El uso de bots ha sido detectado como

una estrategia para inundar tanto a grupos independentistas como constitucionalistas con mensajes polarizantes, incendiarios y violentos. En un estudio que analizó 4 millones de mensajes, se encontró que el 23,6% de los mensajes, retweets y menciones, así como el 38,8% de las respuestas, provinieron de bots automáticos (Stella, Ferrara, & de Domenico, 2018).

El punto de máxima intensidad llegó en la crisis del referéndum ilegal de 2017, especialmente entre septiembre y octubre, cuando las redes pro-Kremlin incrementaron su actividad en favor del proceso en un 2.000%. El día del referéndum, el *hashtag* #Catalanreferendum fue *trending topic* mundial durante 12 horas y las noticias de RT y *Sputnik* se ubicaron entre las fuentes *online* más consultadas, logrando más de 125 millones de visualizaciones. El análisis del tráfico digital reveló que el 84% de las cuentas más activas eran anónimas o gestionadas por bots, con la participación de más de 4.800 cuentas automatizadas, tal como publicó en su informe Iberifier (Badillo, & Arteaga, 2024). Además, Venezuela se estableció como el segundo país de origen en la generación de mensajes a favor del proceso, operando como relevo para amplificar las narrativas rusas. Medios derivados de RT como *Redfish* publicaron y difundieron vídeos en redes, como el titulado *Fighting Franco's Ghost* (peleando contra el fantasma de Franco), en el que se expone a las autoridades españolas con simbología nazi y violencia, frente al pueblo catalán (Redfish, 2017).

En diciembre de 2017, el Centro Estratégico de Comunicación de la OTAN confirmó la existencia de patrones de comunicación de bots en Cataluña semejantes a las campañas de injerencia que se detectaron en Ucrania y Alemania. Estudios posteriores clasificaron la conversación sobre la independencia catalana como una de las cámaras de eco digitales más polarizadas (Colas, 2017), y Twitter cerró cerca de 200 cuentas falsas dedicadas a influir en la opinión pública sobre el *procès* (Badillo, & Arteaga, 2024). En 2019, la revista *Rappler*, de la premio Nóbel de la Paz Maria Ressa, encontraría en Filipinas tres cuentas de Twitter que atribuiría al IRA ruso, que actuaban en su país con comportamiento inusual y muy frecuente, y que anteriormente habían estado emitiendo numerosos mensajes en español agitando el proceso de consulta independentista catalán. Tras esta publicación, Julian Assange, fundador de Wikileaks, sugirió la pertenencia de *Rappler* a la CIA de EE.UU. (Gutierrez, 2018).

El proceso tiene patrones semejantes en lo sucedido posteriormente en torno a grupos independentistas en Escocia, donde cuentas atribuidas a Irán vertieron sobre población específica 250.000 mensajes, 1 millón de reenvíos y 3,2 millones de "me gusta" entre 2021 a 2024 (Linvill, & Warren, 2024).

1.8.2. Granja desde Filipinas como ejemplo de influencia

Es muy complicado, por no decir casi imposible, conocer en su totalidad la estructura de la industria de la influencia, donde se entrelazan la desinformación y el discurso del odio. Pero podemos tener ciertos indicadores al respecto: investigando campañas de desinformación, de cuando los investigadores académicos teníamos acceso gratuito a gran cantidad de datos en la red social Twitter antes de su compra por parte de Elon Musk, pudimos realizar multitud de pruebas para poder ubicar la procedencia de las cuentas que participaban en ellas. Y es que una cosa es lo que una cuenta dice en su perfil sobre su ubicación, y otra muy diferente es su verdadera procedencia. Muchas investigaciones se basan en la información que las cuentas proporcionan sobre su lugar de origen, pero consideramos que esto era una suposición poco confiable. La geolocalización de mensajes inicialmente sólo podía realizarse si el usuario permitía que sus mensajes fueran encontrados con esos datos, pero la mayoría de las cuentas no otorgaban dicho permiso.

Sin embargo, descubrimos que había otra forma de abordar el problema: en lugar de buscar la geolocalización de una cuenta, podíamos invertir el proceso y encontrar los mensajes que provenían de un área específica. A partir de un punto y un radio de kilómetros (no demasiado amplio), se podían extraer todos los mensajes originados en esa zona. Aunque esta técnica no tenía una fiabilidad absoluta, diversos estudios académicos corroboraron que su precisión oscilaba entre el 80% y el 90% en lugares como Europa y Norteamérica, mientras que en regiones de África, Asia o Latinoamérica descendía alrededor del 70%. Aunque no era infalible, este porcentaje era lo suficientemente significativo como para extraer muchas conclusiones. A través de esta técnica, comenzamos a perfilar las procedencias de los mensajes, paralelamente a la recolección de datos general. Un hallazgo recurrente fue que, durante las campañas de desinformación en España en el contexto de la pandemia de covid-19 en 2020, el porcentaje de mensajes provenientes del país se mantenía entre el 35% y el 40%. Este patrón se repitió en diversas campañas desinformativas y bulos que analizábamos durante la pandemia, lo que llevó a extender nuestras áreas de búsqueda.

La tarea fue ardua, ya que se revisaron múltiples puntos y áreas mediante sucesiones de código informático y búsquedas, y durante semanas no se obtuvieron resultados que explicaran el restante 60% de los mensajes en las campañas. Sin embargo, al final, encontramos un punto que comenzó a completar el rompecabezas: un área en Filipinas generaba aproximadamente otro 30% de los mensajes, es decir, la mitad de los mensajes restantes. Esto no significa que los

mensajes realmente provinieron de allí, ya que podría tratarse de un uso de VPN (una especie de puente informático para hacer parecer que los mensajes salían desde esa ubicación, que no dejaría de ser igualmente una información significativa), diversificando así la difusión y esquivando los controles de las plataformas de redes sociales. Además, no debemos olvidar que Filipinas es uno de los países con una de las industrias más desarrolladas en este tipo de actividades. Habíamos detectado con una alta probabilidad, la operativa de una gran granja de mensajes de trolls y bots.

Una vez comprobado que Filipinas era una de nuestras fuentes principales en numerosas campañas, procedimos a intentar "escuchar" todos los mensajes procedentes del lugar, pero el flujo era tan grande que nuestros soportes quedaban en los límites de poder funcionar correctamente. Así, se detectó que durante unos días de junio de 2020 en los que pudimos recoger todo lo procedente del lugar, se obtuvieron un número cercano al medio millón de mensajes en español diarios, con temáticas en su gran mayoría de interés en España, tal como publicamos en un artículo académico al respecto en 2022 (Arce-García, Said-Hung, & Mottareale, 2022). Su clasificación mediante algoritmos a través de los reenvíos producidos da un panorama claro de los grupos a los cuáles se dirigía y cuáles eran las que vertebraban la red encontrada, con casi un millón y medio de mensajes en una muestra recogida de casi tres días. Los catorce principales grupos, que representaban el 72,63% de los mensajes, por tamaño de tráfico de red eran los siguientes:

1. Modelos: especialmente de cuentas que dicen ser chicas jóvenes, muchas de ellas que dicen ser de Mallorca, y que ofrecen multitud de fotos en diferentes posturas sensuales. Un 15,8% del total.
2. Extrema derecha: cuentas que difunden ideología y temas cercanos a estos grupos. El centro de toda la red completa y su intermediación recae principalmente en cuentas de este grupo, con un 13,71% del tráfico de mensajes reenviados.
3. Adolescentes: cuentas que emiten mensajes principalmente dirigidos hacia adolescentes, en los que se introducen muchas imágenes de animales, especialmente gatitos y perritos. 8,13% del tráfico.
4. Izquierda: cuentas seguidoras de los partidos políticos PSOE o Podemos. Se detecta un grupo que, en consulta con uno de los partidos al que supuestamente apoyarían, dicen no conocer de su existencia. 7,47%.
5. Influencers: cuentas que siguen *influencers* variados que hablan de multitud de temas, especialmente para jóvenes. 6,98%.

6. Humor: mensajes que se dedican a difundir chistes en texto, imágenes o memes graciosos. 4,95%.
7. Videojuegos: dedicado a hablar de juegos de diversas consolas y ordenadores. 4,80%.
8. Fuerzas del orden: dirigido especialmente hacia miembros de la Policía Nacional, Guardia Civil y personal del Ejército. 2,34%.
9. Fútbol: relacionado con diversos equipos de fútbol, especialmente primera y segunda división española. 2,17%.
10. Música: comentarios sobre diversos grupos de música. 2,02%.
11. Ciencia: comentarios sobre ciencia, con especial mención a discusiones entre ciencia y aspectos conspiranoicos. 1,69%.
12. K-Pop: es un grupo especialmente dedicado a seguidores de música K-Pop. 1,52%.
13. Televisión, Cine y *streaming*: comentarios sobre programas de televisión, películas o plataformas. 1,48%.
14. Religión: dedicados a hablar de aspectos religiosos. 1,07%.

En las conversaciones se observa un patrón común, como es el intento de ganarse la confianza de otros usuarios de la red. Se pueden notar comentarios muy específicos y locales sobre temas como partidos de fútbol recientes, conciertos de las fiestas de la zona o programas de televisión que se emitieron el día anterior.

Para analizar este perfilado, en la investigación se seleccionaron 20 cuentas de cada uno de los 14 grupos principales identificados, eligiendo al azar entre aquellas que habían emitido al menos tres mensajes durante el periodo de recolección. Se realizó un seguimiento de estas 280 cuentas durante casi tres meses (103 días en total) para observar de qué hablaban. El esquema se confirmó: se trataba de cuentas con bastante actividad, con un promedio de 10,75 mensajes al día de cada una. El análisis de los más de 300.000 mensajes emitidos por estas cuentas reveló que las mismas temáticas se repetían una y otra vez, sobre modelos, ultra-derecha, adolescentes, humor, izquierda, fuerzas del orden, entre otros. La mayoría de los mensajes eran de conversación e introducción al grupo, mientras que un pequeño porcentaje introducía discursos de odio o desinformación, como teorías conspirativas y ataques políticos, dirigidos especialmente a figuras como Pedro Sánchez (Presidente del Gobierno en ese momento), Pablo Iglesias (Vicepresidente), Fernando Simón (portavoz de Sanidad en la lucha contra el covid) y Pablo Casado (presidente del PP, principal partido de la oposición), en el contexto de la pandemia por covid-19. También se incluían ataques contra la inmigración y los okupas.

Los mensajes presentaban una polaridad y sentimientos cuidadosamente medidos para no alertar demasiado a los algoritmos de detección de las plataformas y a otros investigadores. La conexión y relación entre cuentas, que a primera vista parecían desconocidas entre sí y con perfiles tan distintos, podrían expresarse matemáticamente como un efecto denominado de *ultra-small-world*. Este fenómeno en las relaciones humanas demostraría que estaríamos ante una muy alta posibilidad de estar frente a un grupo tan cohesionado que trabajaría en conjunto. Así, se observa un grupo de cientos de miles de mensajes diarios solo en la red social Twitter (ahora X), vertebrado por cuentas de ideología de ultraderecha, que se basan especialmente en muchos medios alternativos de ese mismo perfil ideológico, así como en RT (anteriormente *Russia Today*), que aparece muy conectado a cuentas dirigidas a jóvenes y modelos.

1.8.3. La estructura aumenta

Una vez encontrado un método de trabajo y análisis, que sin ser perfecto si da muy buena aproximación de los movimientos en la red Twitter sobre desinformación y bulos analizados, se aplicó a un caso que ilustra muy bien la forma de difusión y trabajo, un claro trabajo de una técnica denominada *astroturfing* y que se explica en la segunda parte de este libro. Esta técnica consiste, principalmente, en hacer pasar una campaña preparada como un movimiento surgido desde la población en general.

Para ilustrar cómo se lleva a cabo una campaña organizada de este estilo, examinaremos el caso que tuvo lugar en Twitter entre el 14 y el 20 de agosto de 2020, donde se produjo un acoso dirigido hacia el entonces Vicepresidente del Gobierno, Pablo Iglesias, y la Ministra de Igualdad, Irene Montero. Ambos son pareja y pertenecen al partido político Podemos, que compartía en ese momento el Gobierno con el PSOE. Durante ese periodo, Iglesias y Montero estaban disfrutando de unas vacaciones privadas en un pequeño pueblo llamado Felgueras, ubicado en la montaña asturiana, cerca de La Pola (Lena). A través de mensajes en redes sociales, se identificó su ubicación, lo que llevó a que comenzaran a aparecer pintadas en las carreteras que conducían al pueblo, así como acoso en redes sociales y a los hosteleros de la zona. Esta investigación sobre el caso fue publicada en la revista académica portuguesa *Sociología, Problemas e Praticas* en 2022, destacando la forma en que las redes sociales pueden ser utilizadas para coordinar campañas de acoso y desinformación (Arce-García, & Said-Hung, 2022).

Aunque el anuncio público de la presencia de los miembros del Gobierno no se hizo hasta el 15 de agosto de 2020 (14:59) en los diarios asturianos El

Comercio y La Nueva España (16/08/2020 a las 00:56) a través de sus redes, la campaña en redes sociales comenzó el 14 de agosto con mensajes que cuestionaban dónde estaban de vacaciones. Supuestamente, el 15 de agosto de 2020 a las 19:02, la cuenta @_soledad_R, que fue eliminada muy poco después, reveló el lugar exacto de su ubicación: Felgueras. A las 21:05, la cuenta de Podemos de Lena les dio la bienvenida.

Lo interesante de este caso radica en la distribución de los mensajes al realizar un análisis geolocalizado. Aunque casi todas las cuentas en Twitter que participaron afirmaban ser de distintos lugares de España, nuestra recolección de datos estableció que 26.305 mensajes procedían de España (de los cuales 19.389 eran de Madrid, 480 de Asturias, 932 de Valencia, 477 de Málaga y 306 de Sevilla, entre otros). Sin embargo, también se registraron 8.911 mensajes desde Filipinas, 1.544 desde Venezuela, 2.360 desde Estados Unidos, y algunos tuits geolocalizados residualmente en Londres o Nigeria. Aún quedó un 30% de los tuits recabados sin poder ser ubicados por geolocalización. Es importante recordar que se pueden utilizar VPN para desviar la procedencia de los mensajes, pero que aparezcan tantos mensajes desde lugares muy concretos es extremadamente raro y sospechoso. No sólo es relevante la procedencia, sino también la forma de actuar, ya que los mensajes y todos aquellos que los reenvían (retweets) provenían del mismo lugar.

En la siguiente tabla se pueden apreciar ejemplos de diversos mensajes, aunque hay muchos más, todos con un patrón de funcionamiento similar:

Lugar	Mensaje	Tuits	Inicio	Final
Londres	"Joe!!! Ya ta Asturias llenándose de esta mierduza roja!! Iglesias y Montero de vacaciones en la casa asturiana del líder del PCE en plena crisis por la 'caja B'"	3	16/08/20 14:36	16/08/20 14:50
Filipinas	"Qué vergüenza, como asturiana me siento avergonzada al oír en la TV que Pablo Iglesias e Irene Montero han tenido que abandonar sus vacaciones en Asturias por su seguridad. Hubo gente que fue a insultarlos y amenazarles a la casa donde se alojaban."	474	17/08/20 15:11	19/08/20 12:35

Lugar	Mensaje	Tuits	Inicio	Final
Estados Unidos	“Pablo Iglesias e Irene Montero se han tenido que marchar, por seguridad, del domicilio donde iban a pasar 1 semana de vacaciones con sus hijos, en Asturias. Hay que dar las gracias a un periódico de la zona que ha puesto la dirección para que la extrema derecha sigan acosándoles”	945	17/08/20 15:17	20/08/20 05:39
Filipinas	“Son "queridos" allá donde van. Quien siembra vientos recoge... Montero e Iglesias abandonan la casa donde veraneaban en Asturias tras sufrir acoso, insultos y amenazas”	82	17/08/20 17:08	18/08/20 14:30
Venezuela	“Pablo Iglesias e Irene Montero han tenido que abandonar la casa de Asturias por el "acoso" que sufrían por los vecinos. Quien a hierro mata, a hierro muere. Que se jodan.”	212	17/08/20 17:17	19/08/20 02:28
Filipinas	“La escolta de Iglesias y Montero no vio acoso en su estancia en Asturias. #SiguemeYTeSigo https://t.co/6g0l8iyp1d”	50	20/08/20 00:11	20/08/20 18:55
Estados Unidos	“¿Se lo inventaron? La escolta de Iglesias y Montero no vio acoso en su estancia en Asturias | Diario Sur https://t.co/bKjox6Rgmy”	22	20/08/20 09:45	20/08/20 19:10
Nigeria	“@IreneMontero Se os ve el plumero, estéis buscando excusas para provocar enfrentamientos civiles, para esa “ruptura” que tanto añoráis y para desviar la atención sobre vuestras corruptelas.”	1	20/08/20 15:36	20/08/20 15:36

Tabla 1. Mensajes y Retweets enviados desde distintos lugares

En estas conversaciones se puede observar cómo, desde los mismos lugares, surgen discusiones y enfrentamientos entre grupos de cuentas. Estos grupos parecen estar utilizando ataques de falsa bandera, donde aparentemente se enfrentan a la parte contraria, generando polaridad y confrontación. Los mensajes se escriben de manera sucesiva desde los mismos lugares geolocalizados,

creando *hashtags* dedicados como #SiguemeYTeSigo (entre los atacantes) y #LaFascistaDeLaSole (entre los defensores). Este último *hashtag* se refiere a "Sole", la supuesta autora de la cuenta que divulgó la ubicación de las vacaciones de Pablo Iglesias e Irene Montero. Como se mencionó anteriormente, esta cuenta fue eliminada poco después de emitir su mensaje, que incluía una foto de una mujer que aparecía en otras imágenes, siempre en primer plano y ubicadas cerca del campo de fútbol Carlos Tartiere, del Real Oviedo. Sus fotos también aparecieron en otras cuentas con nombres diferentes, como @_soledad_Ro, que también fue eliminada.

Este enfrentamiento en redes es un claro ejemplo del funcionamiento de la generación artificial de discusiones y polarización, pero es que además esta historia tiene un apéndice muy ilustrativo: el salto de la red a la vida real, conocido como el salto *online* al *offline*. El 18 de agosto, una cuenta señaló que la autora que ubicó el lugar, "la Sole", era la propietaria de una frutería en el mercado de La Felguera (un nombre similar a Felgueras, pero que se encuentra a 26 km de distancia). Esto provocó que recibiera acoso telefónico y en redes sociales, lo que llevó a los propietarios de la tienda a intervenir y pedir que cesara dicho acoso a los líderes políticos de Podemos e Izquierda Unida de la localidad.

Lugar	Mensaje	Tuits	Inicio	Final
Estados Unidos	"A ver, zoquete. Qué cojones tiene que ver una frutería de La Felguera que se llama Sol con la tía esta que vive en Felgueras? Tenéis lo justo para no cagaros encima."	2	18/08/20 11:50	18/18/20 12:12
Filipinas	"Está circulando por las redes un teléfono de Frutería Sol de La Felguera (municipio de Langreo), que no tiene nada que ver con la dichosa Sol de Felgueres (municipio Lena). Por favor reenviar éste mensaje porque a la mujer le están quemando el teléfono #LaFascistaDeLaSole"	58	18/08/20 13:15	19/08/20 12:46

Tabla 2. Mensajes y Retweets enviados desde distintos lugares

De los 558 mensajes de Twitter involucrados en este apéndice de la historia, 135 procedían de España (127 desde Asturias), 57 desde Filipinas y 2 desde Estados Unidos, mientras que el resto no pudo ser geolocalizado. En todas las fases

de esta campaña, los algoritmos de análisis determinaron que aproximadamente el 52% de las cuentas que participaron eran bots. Además, la gran mayoría de estas cuentas mostraban una actividad muy alta en redes sociales todos los días, con un perfil muy característico.

Este ejemplo ilustra muy bien el funcionamiento de muchas campañas que se pudieron analizar en el periodo en el que los investigadores podíamos estudiar la red social X, que hoy ofrece muchos menos datos, y bajo grandes sumas de dinero de pago. Pero el tiempo en que se pudo trabajar hasta 2023, se pudieron detectar mensajes en gallego desde Macedonia del Norte, en catalán desde Bélgica y países de oriente medio, pero especialmente desde Filipinas, Venezuela, Nigeria, Estados Unidos o Inglaterra hacia muy diversas campañas de cualquier tipo y en cualquier sentido claramente del ámbito español. Cabe insistir en que la procedencia aparecieran desde dichos países no significa que fueran realmente desde allí, pero expresarían de todas maneras una necesidad de ocultar la procedencia real de unos mensajes.

Nuestros estudios dentro del proyecto Hatemedia (2025) nos han llevado a multitud de datos que vienen a corroborar la gran sospecha de estar ante un gran entramado organizado. A través de la recogida de más de 10 millones de mensajes en plataformas como X/Twitter, Facebook o los foros de los principales diarios de noticias españoles, a partir de contestaciones de distintos usuarios comentando noticias de dichos periódicos digitales de distintas tendencias ideológicas, se encontró que el 51,45% de los mensajes contenían odio en distintos tipos de intensidad (no solo ofensivos o amenazantes desde un punto de vista legal), percibiendo que lo que se emplea son los mensajes de odio de intensidad lo suficientemente alto como para causar impresión, pero ser lo suficientemente bajo como para tener problemas legales o ser eliminado fácilmente por parte de plataformas. El análisis que publicamos en la revista *New Media & Society* (Arce-García, Said-Hung, & Montero-Díaz, 2024) viene a demostrar como multitud de cuentas, que denominamos nano- o micro-influencers (por tener unos pocos cientos de seguidores únicamente y hacerse parecer como cuentas de personas normales y corrientes), con un comportamiento estructurado y coordinado actúan propagando odio mediante técnicas de propagación *astroturfing*. Este odio de baja intensidad, pero continuo, también ha sido detectado en estudios en torno a la propagación de odio en Inglaterra alrededor de cuentas de ideología de extrema derecha (Vidgen, 2019; Williams, 2021).

El seguimiento de las cuentas que emiten odio, como decía anteriormente detectadas a partir de réplicas a informaciones de distintos diarios españoles,

vienen a reflejar distintos puntos destacables en nuestros estudios, especialmente en X:

- Emiten a horas y días de la semana coordinados. Incluso ciertos temas se producen más en determinados días de la semana que en otros: cada día hay un tipo de odio, y es más elevado de martes a jueves, bajando mucho los fines de semana.
- Sus mensajes de odio y de desinformación son siempre en torno al 10% del total de sus mensajes, con lo que la mayor parte del tiempo intentan pasar desapercibidos y contactar con otras cuentas a través de distintos temas.
- Sus estilos de escritura son semejantes, tienen patrones estilométricos de estructura semántica que permite detectar que de 1.000 cuentas seguidas durante un mes, había sobre un 70% de ellas que tenían un patrón semejante, por lo que estarían bajo unas mismas líneas editoriales o guión predefinido previo, que se modifican un poco para parecer distintas. El estilo de escritura en los mensajes está ajustado en un 91% de los casos para comprensión de un nivel educativo en torno a primaria o primer año de secundaria (12-13 años), solo un 2,5% tiene un estilo de escritura de bachiller o superior. El uso emocional es elevado, especialmente de miedo y asco.
- Sus mensajes y seguimientos tiene grandes conexiones, como por ejemplo que durante todos los mensajes y reenvíos del mes de enero de 2021, unas 1.000 cuentas identificadas por emitir odio en distintos lugares y tiempos, tuvieran todas en común seguir y apoyar en su red a la cuenta @ToniCanto1 (cuenta de un actor y político por entonces del partido Ciudadanos en Valencia, poco antes de dimitir y pasar a la candidatura del PP por Madrid) como centro de su red y, en menor medida, a otras cuentas de difusión en su mayoría de ideología de extrema derecha.
- Entre 2.243 cuentas odiadoras identificadas en otro estudio del proyecto, igualmente a partir de comentarios en diarios españoles, se aprecia que en su historial de mensajes a lo largo de febrero de 2021, en torno al 70% de ellas procedían de grupos de ideología de extrema derecha, un 15% de cuentas de ideología de extrema izquierda, un 8% de cuentas de procedencia o comentarios de Venezuela y en menor medida Colombia, Perú y otros países de centroamérica con conexiones con medios rusos como RT, y en los restantes aparecen grupos pro-Palestina o manga japonés.

> Entre las cuentas, se aprecia que, cuando no emiten odio (90% de sus mensajes no emiten odio), hablan de deporte, música o animales, pero lo que destacan son sus valores de cohesión de red (longitudes de red, modularidad, etc.) y aparente coordinación y dependencia entre sí en cada grupo identificado. Su actividad es bastante elevada, con medianas de entre 11,84 a 35,63 mensajes diarios emitidos por cada cuenta de cada grupo identificado.

Todo ello dibuja un escenario en el que circulan mensajes de carga de odio de baja-media intensidad, emitidos por usuarios cuyo comportamiento, lejos de ser aleatorio, sigue pautas reconocibles en la red social X. El hecho de que algo más de un par de miles de cuentas, supuestamente ajenas entre sí, muestran semejanzas tan marcadas en su manera de intervenir resulta difícil de atribuir al azar, y puede compararse con la imagen de un salón de baile en el que todas las parejas repiten exactamente los mismos pasos al mismo tiempo. Una coincidencia tan precisa invita a pensar en algún tipo de coordinación o aprendizaje compartido de pautas de comportamiento. De este modo, los contenidos hostiles se entrelazan con mensajes aparentemente neutros o informativos, contribuyendo a crear un clima discursivo en el que el odio se percibe como una reacción más dentro del flujo cotidiano de comentarios.

A partir de estas dinámicas puede presumirse la existencia de una red principal de difusión de odio en España articulada en torno a los medios informativos digitales, cuya actividad se suma a la presencia de otras redes más pequeñas, en particular de determinadas áreas de Latinoamérica.

1.8.4. La propaganda pro-Kremlin

Aunque los datos anteriores muestran que en España principalmente se estaría ante una red que se vertebra principalmente entre cuentas de difusión de extrema-derecha, existe a distancia una red de influencia por cantidad y alcance, pero no por ella menos importante. Incluso en ocasiones, ambos grupos tienen sinergias o campañas en las que confluyen. La red pro-Kremlin de desinformación en España suele describirse como un sistema en tres capas que trabajan en cascada: arriba, los medios y proyectos impulsados o controlados por el Estado; en el medio, los canales "embudo" en español que traducen, reinterpretan y empaquetan esos mensajes; abajo, una constelación de cuentas y comunidades locales que los difunden dentro del debate político español. En la cúspide están marcas como el entramado de webs tipo *Pravda*, junto con operaciones encubiertas que crean páginas que imitan a medios occidentales para publicar piezas

manipuladas sobre España y la UE. Esa producción se orienta explícitamente a debilitar el apoyo a Ucrania, atacar a la OTAN y presentar a la Unión Europea como un proyecto fallido o decadente, y es la materia prima informativa que luego circula en castellano con apariencia de análisis neutral o periodismo independiente (Maldita, 2025a; Thomas, & França, 2025; González, 2025).

La segunda capa la forman canales en español que actúan como traductores y amplificadores de esa materia prima, sobre todo en la red social Telegram. Investigaciones recientes han identificado una red de decenas de canales impulsados o promocionados que, en un solo año, llegaron a publicar decenas de miles de mensajes en castellano dirigidos tanto a audiencias españolas como latinoamericanas. En estos espacios se reciclan noticias de agencias y televisiones rusas, se añaden comentarios políticos alineados con el Kremlin y se conectan temas globales con polémicas internas españolas, desde el coste de la vida hasta las protestas del campo o la política migratoria. Estos canales se presentan a menudo como proyectos "alternativos", críticos con los grandes medios, pero su pauta de publicación y el origen de los contenidos apuntan a una función clara de embudo entre la esfera mediática rusa y el usuario hispanohablante (Monitor Disinfo, 2023; Biescas, Allen, & Hernández, 2024)

En la tercera capa aparecen cientos de cuentas y comunidades que operan ya dentro del ecosistema político y social español, muchas veces sin vínculos formales con Rusia, pero sí con una afinidad ideológica que las lleva a amplificar esos mensajes de forma selectiva. Aquí caben pequeños canales de Telegram de provincias, grupos de WhatsApp, perfiles de X o páginas de Facebook que combinan críticas legítimas al gobierno con teorías conspirativas o bulos que proceden, unas horas antes, de los nodos pro-Kremlin. En crisis como la DANA que golpeó la Comunidad Valenciana o los apagones eléctricos que afectaron a España y Portugal, se ha visto con claridad el ciclo: primero surgen piezas manipuladas o suplantaciones de medios extranjeros en el circuito ligado a Moscú, luego pasan por los canales embudo en español y, finalmente, los comparten cuentas locales que las usan como munición para denunciar el supuesto colapso del Estado o la traición a los intereses nacionales (Terrero Carrobles, 2025; Artuch, 2025).

En cuanto a la financiación, las investigaciones apuntan a un modelo mixto, donde los primeros recibirían financiación directa o indirecta del Estado ruso, que sostiene medios, agencias y proyectos de influencia orientados a Europa. En medio, canales que se presentan como independientes pero que piden donaciones mediante plataformas de pago occidentales y, en algunos casos, remiten a cuentas bancarias o monederos de criptomonedas vinculados a entidades rusas sancionadas. Abajo, una militancia más bien voluntarista, formada por

activistas y creadores de contenido que monetizan su actividad a través de pequeñas donaciones, publicidad y visibilidad dentro de sus nichos ideológicos (Fuente Cobo, 2025; Maldita, 2025b; Borràs Rius, 2025).

En la intersección entre esa red propagandística y el terreno puramente técnico aparece *NoName057(16)*, el grupo de *hackers* pro-Kremlin que ha protagonizado buena parte de los ciberataques contra instituciones españolas desde 2022, entre otros países europeos. El grupo llegó a tener hasta 4.000 simpatizantes activos y en julio de 2025 sufrió detenciones en varios países de Europa en la operación *Eastwood*, dos de ellas en España (Arias, 2025) y buscado un español refugiado en Rusia (López-Fonseca, 2025). Este colectivo se organizaba en torno a canales de mensajería en Telegram donde se anuncian objetivos, se reclutan voluntarios y se distribuye la herramienta DDoSia, que permitía lanzar ataques coordinados contra Ministerios, parlamentos autonómicos, ayuntamientos, medios de comunicación o procesos electorales, a cambio de micropagos en criptomonedas a quienes participan con más intensidad. Esos golpes no se quedan en el plano técnico, pues cada ataque iba acompañado de mensajes de propaganda que presumen del impacto, se burlaban de las autoridades españolas y lo encuadran como castigo por apoyar a Ucrania, de modo que la propia actividad de *NoName* servía como combustible simbólico para la red de desinformación pro-Kremlin, que presentaba los fallos puntuales de servicios públicos como prueba de una supuesta debilidad estructural del Estado (Monitor Disinfo, 2025; Euronews, 2025). Desde agosto de 2025 se produjo una reactivación de los ataques dirigidos a entidades públicas españolas, en represalia por las detenciones y bajo las operaciones #timetoretribution, #fuckingeastwood y #opSpain. Entre los objetivos recientes se encuentran ayuntamientos, portales de transporte, organismos públicos y empresas españolas. El grupo ahora cuenta con el respaldo de colectivos aliados como *Z-pentest* y *Mr.Hamza*, buscando demostrar que su capacidad operativa sigue intacta. En octubre de 2025 intentaron nuevos ataques contra varias administraciones, aunque con escaso éxito (Méndez, 2025, Balmaceda, 2025).

1.9. RESUMEN DE LA PRIMERA PARTE

Se expone a continuación un resumen cronológico de lo más destacado expuesto en esta primera parte:

Antigüedad y Siglo XIX

- Siglo XII a.C.: Ramsés II ordena la construcción de Abu Simbel, utilizando grabados falsos de hititas ejecutados para transformar el empate de la batalla de Qadesh en una victoria personal.

- Siglos VI-V a.C.: Sun Tzu publica "El Arte de la Guerra", estableciendo que controlar la percepción del enemigo permite ganar sin enfrentamiento directo.
- 1898: Se graba en Brooklyn, Estados Unidos, la explosión de una maqueta del fuerte de La Habana para incitar a la guerra contra España.

1917 - 1950

- 1917: Se establece la Cheká en Rusia para gestionar la propaganda y eliminar cualquier desviación de la línea oficial post-revolucionaria.
- 1919: Edward Bernays abre su primera consultoría de "consejero en relaciones públicas" tras trabajar en el Comité de Información Pública de Woodrow Wilson.
- 1921-1927: La Cheká en Rusia ejecuta la Operación Confianza, creando una falsa organización monárquica para atraer y neutralizar a opositores zaristas.
- 1925 (3 de marzo): El Parlamento británico debate oficialmente por primera vez la expansión de la "propaganda subversiva" en el Imperio.
- 1928: Bernays publica su obra fundamental "Propaganda", teorizando sobre el "gobierno invisible" que manipula a las masas.
- 1929: Bernays organiza el desfile "Antorchas de Libertad" en la Quinta Avenida, vinculando el tabaco con el voto femenino.
- 1938: Orson Welles emite el radioteatro de "La Guerra de los Mundos", provocando que más de un millón de personas creyeran en una invasión real.
- 1939-1945 Segunda Guerra Mundial: El agente británico Sefton Delmer dirige la emisora clandestina Gustav Siegfried Eins (GS1) para emitir "propaganda negra" contra los nazis.

1950 - 2000

- 1951-1954: Bernays diseña para la United Fruit Company la campaña de desinformación que justifica el golpe de la CIA en Guatemala contra Jacobo Árbenz.
- 1960: El KGB distribuye panfletos falsos atribuidos al Ku Klux Klan entre delegaciones de la ONU para desprestigiar a EE. UU. ante países africanos y asiáticos.
- Años 80: El KGB ejecuta la Operación Denver, logrando que comunidades conspiranoicas crean que el SIDA fue creado en un laboratorio militar estadounidense.

- 1990: Newt Gingrich publica el memorando "Language: A key mechanism of control", listando palabras clave (traidor, radical, corrupto) para polarizar el discurso político.
- 1990: Se funda Strategic Communication Laboratories (SCL Group), empresa matriz de la futura Cambridge Analytica.
- 1991 (19 de agosto): Durante el intento de golpe de Estado en Rusia, la red Relcom transmite 46.000 noticias independientes para coordinar la resistencia frente a la censura oficial.
- 1991: China eleva su oficina de Propaganda Exterior a nivel ministerial.
- 1996: El ejército de EE. UU. oficializa el concepto de "Dominación Informativa" (Information Dominance) en su doctrina militar.
- 1997: Alexander Dugin publica "Los fundamentos de la geopolítica", proponiendo el uso de la desestabilización discursiva como arma geopolítica.
- 1997: El coronel ruso Sergei Komov publica los 11 principios del "control reflexivo" para influir en las decisiones del adversario.
- 1998: Se detecta MoonLight Maze, la primera gran ciberoperación rusa contra sistemas militares críticos de EE. UU.

2001 - 2013

- 2001 (11S): Los atentados de Nueva York impulsan el nacimiento de la "economía de la vigilancia" y el uso masivo del Big Data para seguridad nacional.
- 2003: Se fundan el foro 4chan y la empresa Palantir Technologies.
- 2004: China activa la "Armada de los Cincuenta Céntimos" (Wumao) para saturar los foros digitales con mensajes favorables al régimen.
- 2006: Se documentan las primeras tácticas de manipulación en blogs y foros de oposición durante las elecciones de México.
- 2007: Se funda Breitbart News, que se convertiría en el nodo central de la *alt-right* y el trumpismo.
- 2007: Estonia sufre un ataque DDoS masivo que aísla digitalmente al país tras el traslado de una estatua soviética.
- 2008: Barack Obama integra por primera vez las redes sociales y el microtargeting de datos en su estrategia electoral.
- 2008-2009: Israel lanza la operación "Plomo Fundido", estableciendo la *Hasbara* digital para contrarrestar críticas internacionales en YouTube y Twitter.
- 2010-2012: Las Primaveras Árabes, el 15M y Occupy Wall Street marcan el auge de la tecnopolítica y la protesta digital global.

- 2011: Tras denuncias de fraude electoral en Rusia, se crean ejércitos de trolls para controlar la narrativa en VKontakte y Odnoklassniki.
- 2012: Campañas en México utilizan ejércitos digitales para posicionar hashtags.
- 2013: Se fundan el Internet Research Agency (IRA) en San Petersburgo y el foro 8chan.

2014 - 2019

- 2014: El *Gamergate* estalla como una campaña de acoso masivo contra desarrolladoras de videojuegos, sirviendo de laboratorio para la guerra cultural de la alt-right.
- 2014: Cambridge Analytica extrae datos de 87 millones de usuarios de Facebook.
- 2014 (Julio): Rusia despliega una campaña de desinformación tras el derribo del vuelo MH17, inventando la figura del controlador aéreo falso "Carlos Spainbuca".
- 2014: SCL Group realiza en Trinidad y Tobago el "Proyecto Trinité" para predecir comportamientos electorales con flujos de datos nacionales.
- 2014-2020: Se desarrolla la campaña rusa "Infección Secundaria", distribuyendo 2.500 contenidos falsos en 7 idiomas a través de 300 plataformas.
- 2015: El empresario español Javier Pérez Dolset distribuye el software "Social Baton" (diseñado en Rusia) para realizar campañas políticas.
- 2016: La campaña Vote Leave en el Brexit en Reino Unido utiliza a AggregateIQ (filial de Cambridge Analytica) para enviar mil millones de mensajes personalizados.
- 2016: La ciudad de Veles (Macedonia del Norte) se convierte en un centro de producción de noticias falsas pro-Trump con fines puramente económicos.
- 2016 (Diciembre): Un hombre armado asalta una pizzería en Washington D.C. convencido por el bulo del Pizzagate.
- 2016: Proyecto Álamo para las elecciones de Estados Unidos por parte de Cambridge Analytica.
- 2017: Surge el movimiento QAnon en 4chan y el escándalo de los MacronLeaks en Francia.
- 2017 (Octubre): Durante el referéndum en Cataluña, la actividad de las redes pro-Kremlin aumenta un 2.000%, con un 84% de cuentas gestionadas por bots.

- 2018: Se disuelven Cambridge Analytica y SCL Group tras las filtraciones de Christopher Wylie.
- 2018: Jair Bolsonaro gana las elecciones en Brasil apoyado por "máquinas de noticias falsas" en redes sociales.
- 2019 (15 de marzo): Un supremacista blanco atenta contra una mezquita en Christchurch (Nueva Zelanda), retransmitiéndolo en vivo tras anunciarlo en 8chan.

2020 - Presente y Futuro

- 2020: Durante la pandemia, colectivos recaudan miles de euros para financiar campañas negacionistas.
- 2021: Se produce el asalto al Capitolio en EE. UU. (6 de enero) y el caso del "Barçagate" en España.
- 2022: Rusia inicia la invasión de Ucrania y despliega la campaña Doppelgänger, clonando dominios de medios europeos para difundir desinformación.
- 2023: Sale a la luz "Team Jorge", grupo mercenario israelí que ofrece servicios de hackeo y desinformación en elecciones de 30 países.
- 2024: Durante la DANA en Valencia, redes pro-Kremlin y chinas amplifican bulos para atacar a las instituciones españolas, incluyendo suplantaciones de ONGs.
- 2024 (Noviembre): El Tribunal Constitucional de Rumania anula la primera vuelta electoral por injerencias masivas de cuentas de TikTok vinculadas a redes rusas.
- 2025 (Diciembre): EE. UU. deniega el visado a reguladores europeos acusándolos de censura digital.
- 2026 (Enero): Rusia prevé completar el control absoluto de su red nacional (RuNet) para posibilitar la desconexión total del internet global.

2. ¿Cómo se crea una campaña de desinformación y odio?

Como se explicó en la primera parte de este libro, la desinformación y los discursos de odio no son fenómenos nuevos en la historia. Sin embargo, hay una diferencia crucial en la actualidad: en los últimos años han cobrado fuerza gracias a las nuevas tecnologías y a los hallazgos de estudios sociológicos y psicológicos que amplifican su impacto. Hoy en día, el uso de redes sociales e internet permite que las personas generen miles de datos cada día, principalmente a través de sus teléfonos móviles.

Para ilustrar esto, decidí instalar una VPN (red privada virtual) en mi teléfono con el fin de ocultar mi dirección IP y navegar de manera más privada y segura. Una de las características que ofrece esta VPN es la capacidad de registrar los intentos de las aplicaciones y programas en mi dispositivo para recopilar información sobre mí y enviarla a empresas que gestionan esos datos. No tengo aplicaciones inusuales: uso de aplicaciones de transporte, algunos juegos sencillos, música, idiomas y redes sociales. Sin embargo, en la última semana, 22 aplicaciones han intentado acceder a mis datos, con un total de 19.738 intentos de rastreo.

¿Y qué buscan exactamente en mi teléfono? Información como mi dirección, ubicación geográfica, identificador comercial, zona horaria, memoria del dispositivo, densidad de la pantalla, idioma, tiempo de arranque al encender el teléfono, orientación del móvil, uso de auriculares, nivel de batería, operador de red, volumen, mi correo electrónico, entre muchos otros datos. Pero, ¿para qué necesitan toda esta información? Para clasificarme dentro de un grupo de usuarios, en un clúster que un algoritmo de asociación determinará. A partir de ahí, podrán deducir mi nivel adquisitivo, mis patrones de movimiento, mis relaciones sociales y si llevo una vida tranquila o si estoy siempre en movimiento. Cada dato cuenta para clasificarnos y, en última instancia, para vendernos productos, ya sean materiales o inmateriales como ideas o gustos.

Hoy en día, proporcionamos una cantidad asombrosa de datos de manera casi inconsciente, y además, compartimos voluntariamente información personal,

como fotos de nuestras vacaciones, fiestas, amigos y logros laborales de los que nos sentimos orgullosos. Al hacerlo, estamos entregando una gran cantidad de información que permite crear perfiles comerciales basados en nuestras características. Estas bases de datos se han convertido en algunos de los activos más valiosos del mundo, generando más ingresos que muchos sectores productivos.

Un buen ejemplo de este fenómeno es el caso de Cambridge Analytica, que ya apareció en un capítulo anterior. La empresa afirmaba ser capaz de predecir, con una fiabilidad superior al 85%, rasgos tan diversos como el género, la orientación sexual, las inclinaciones políticas o incluso si una persona era hijo/a de padres divorciados, a partir de sólo sesenta y ocho "me gusta" en Facebook. Su equipo llegó a elaborar casi cincuenta tipologías distintas de personalidad basadas en patrones de comportamiento digital. Si con una cantidad tan limitada de datos es posible obtener diagnósticos tan precisos, resulta fácil imaginar el alcance que puede tener la explotación de la ingente información que generamos cada día. Navegar por internet o participar en redes sociales equivale, en cierta forma, a caminar por la calle relatando a desconocidos aspectos íntimos de nuestra vida tales como quiénes son nuestros amigos, nuestros hijos o a dónde hemos ido de vacaciones. Sin embargo, millones de personas comparten ese tipo de detalles sin reparo alguno, movidas por una necesidad profunda de conexión, de validación o de pertenencia en una sociedad donde lo digital ha sustituido en buena parte al contacto humano presencial.

Esa exposición voluntaria ha sido convertida en oro por la industria del marketing. Las empresas ya no necesitan invertir en campañas masivas que alcanzan a públicos irrelevantes para sus productos. La publicidad, sostenida por algoritmos de segmentación, se ha vuelto quirúrgica, pues una compañía de videojuegos de acción no desperdiciará recursos intentando atraer a personas mayores, del mismo modo que una marca de cosmética masculina no se dirigirá a quienes jamás han mostrado interés por ese mercado. Cada anuncio aparece ante los ojos de potenciales compradores cuya probabilidad de conversión ha sido calculada con precisión matemática. En este nuevo ecosistema, los datos se han convertido en la materia prima esencial de la economía digital. El verdadero negocio de muchas plataformas no consiste en vender programas o aplicaciones, sino en extraer, analizar y comercializar la información que los usuarios generan a cambio de servicios aparentemente gratuitos. En esa transacción invisible, los datos personales se transforman en la moneda con la que pagamos nuestra experiencia en línea, alimentando un sistema donde el conocimiento sobre nuestras vidas se convierte en la clave de la rentabilidad empresarial.

Sin embargo, centrémonos en el tema que nos ocupa, que no deja de ser aplicar el marketing dentro de la industria de la desinformación, el odio y la manipulación de la opinión pública. El objetivo de esta sección es desentrañar cómo funciona este tipo de marketing que, en esencia, es una venta de ideas destinadas a alterar la percepción del público. Este fenómeno se relaciona con lo que en el ámbito de la comunicación política se conoce como el desplazamiento de la ventana de Overton. Este concepto, desarrollado por Joseph P. Overton (1960-2003), un miembro del *Mackinac Center for Public Policy* en Michigan, comenzó a delinear en los años 90 un espectro de ideas que serían consideradas aceptables o inaceptables por la sociedad. Tras su fallecimiento, Joseph Lehman continuó su trabajo y le dio su nombre en reconocimiento a su contribución.

La teoría establece que la opinión pública determina qué ideas y acciones son vistas como aceptables y populares, mientras que otras son consideradas inaceptables o impensables. Cualquier propuesta que se encuentre fuera de esta "ventana" de ideas admisibles será rápidamente desechada. Por ello, los políticos dedican mucho tiempo a identificar la ubicación de esta ventana para alinearse con ella y conectar con la población. Sin embargo, el marketing político puede ser utilizado justamente para desplazar esa ventana de Overton. Este movimiento ocurriría de manera gradual, sin grandes saltos, permitiendo que ciertos aspectos que antes eran inaceptables comiencen a ser aceptados, mientras que otros que solían ser aceptables caerían en desuso. Aunque esta teoría ha ganado popularidad en la actualidad, también ha sido objeto de críticas, siendo acusada de ser una teoría de conspiración destinada a manipular a la sociedad (especialmente por aquellos que podrían estar beneficiándose de su aplicación).

En el terreno de la seguridad informática se asume desde hace años que resulta más sencillo engañar a una persona con acceso privilegiado que vulnerar directamente un sistema bien protegido: el eslabón débil no suele ser el cortafuegos, sino el usuario. A partir de esta constatación, organismos como la Unión Europea, la OTAN o distintos gobiernos, entre ellos el español, han comenzado a utilizar marcos analíticos específicos para comprender mejor cómo se explota ese punto débil humano y cómo se pueden anticipar y neutralizar este tipo de ataques. Buena parte de estos trabajos se apoyan en un modelo procedente del mundo militar y de la ciberseguridad, conocido como *kill chain*, que ha sido progresivamente adaptado al estudio de fenómenos sociales y comunicativos. Este enfoque propone descomponer cualquier ataque en una secuencia de fases, de modo que se puedan analizar con detalle los pasos que seguiría un actor hostil para alcanzar su objetivo. Aplicado a la desinformación o a los dis-

cursos de odio, implica describir, fase por fase, cómo se concibe, planifica, despliega y amplifica una campaña hasta que logra instalar sus narrativas en el debate público o en comunidades concretas.

En esta sección el propósito es desentrañar cómo se diseñan y ejecutan las campañas de manipulación, desinformación y difusión de discursos de odio, y por qué resultan tan eficaces a la hora de explotar nuestras vulnerabilidades individuales y colectivas. Nos detendremos en un ámbito en plena expansión, la ingeniería social, que se ha convertido en una herramienta central para engañar y dirigir el comportamiento de las personas hasta el punto de comprometer tanto su seguridad personal como la de las instituciones y comunidades a las que pertenecen.

La hipótesis de partida que guía este capítulo es sencilla, pero poderosa, pues cuanto mejor se conozca la forma de pensar y actuar de quienes impulsan estas campañas, más fácil será detectarlas a tiempo y diseñar respuestas eficaces. Comprender la cadena completa de acciones del atacante (desde la selección de objetivos hasta la elección de canales, mensajes y tácticas de ingeniería social) no solo permite interpretar lo que está ocurriendo, sino también intervenir en los puntos críticos donde la ofensiva resulta más vulnerable.

A partir de este enfoque, se establecerá un modelo conocido como TTP (Tácticas, Técnicas y Procedimientos) con los aspectos fundamentales que un proceso de *kill chain* debería seguir: las tácticas se centran en el análisis de cómo los atacantes establecen sus objetivos, ideas de acción y planificación; las técnicas son los métodos que se desarrollan para alcanzar los objetivos fijados en las tácticas; y por último, los procedimientos detallan cómo aplicar esas técnicas de manera efectiva.

Existen diversos modelos para abordar la manipulación y desinformación, pero uno de los más utilizados en la actualidad es el marco DISARM. Aunque es relativamente reciente, desarrollado a partir de 2018 y basado en otros modelos TTP como el MITRE ATT&CK (n.d.) de ciberseguridad, DISARM ha ganado popularidad rápidamente entre los especialistas, profesionales y académicos. Esta iniciativa fue creada por la *DISARM Foundation* (n.d.), que reúne a varias instituciones y universidades de Estados Unidos y Europa. En Europa, el actor principal detrás de DISARM es *Alliance4Europe* (n.d.), una organización sin ánimo de lucro que busca defender la democracia en el continente frente a las amenazas de desinformación, y que cuenta con la participación de ONG, así como de entidades como el Parlamento y la Comisión Europea, y fundaciones dedicadas a la cultura democrática. Sin embargo, algunos grupos conspiranoicos, de extrema

derecha o pro-Kremlin han acusado a *Alliance4Europe* y su metodología de ser herramientas de censura y manipulación.

Para describir el proceso de manipulación y desinformación seguiremos, por tanto, en gran medida el modelo *DISARM Framework* (n.d.). Este enfoque es utilizado por numerosos países, así como por la Unión Europea y la OTAN, especialmente en relación con los ataques denominados de tipo FIMI (*Foreign Information Manipulation and Interference*). Este término se refiere a las operaciones de manipulación e interferencia extranjera en el ámbito de la información y se emplea actualmente para definir y analizar las amenazas desinformativas que llegan a la Unión Europea desde el exterior, pero puede ser empleado ante cualquier tipo de campaña desinformativa y/o odio organizada, ya que su patrón de comportamiento es similar, con mayor o menor grado de sofisticación. Justamente el grupo de Acción Exterior de la UE publicó el marco de respuesta a las amenazas FIMI para organizar y dar concepto a todos estos procesos, especialmente provenientes de países como Rusia o China (EUvsDisinfo, 2025). Obviamente no todas las campañas ni empresas u organismos que las difunden tienen tal grado de sofisticación y medios como para llevar a cabo cada uno de estos procesos que se van a describir, pero en mayor o menor medida, siguen estos patrones de funcionamiento a la hora de trabajar.

Igualmente, para apoyar en la explicación de las técnicas, se hará uso del trabajo realizado por el Foro contra las Campañas de Desinformación, organizado por el Ministerio de Presidencia y el Departamento de Seguridad Nacional de España en sus iniciativas de 2024, en su primer capítulo (Arce-García, & et al., 2025). Dicho grupo, que tuve el honor de codirigir junto a la Dra. Leticia Rodríguez Fernández (Universidad de Cádiz) y un miembro del Dpto. de Seguridad Nacional, así como los restantes autores: Mª José Establés Heras (U. Castilla La Mancha), David García Marín (U. Rey Juan Carlos), Beatriz Marín García (European External Action Service-EEAS), Virginia Martín Jiménez (U. Valladolid), Concha Pérez Curiel (U. Sevilla), Elías Said Hung (U. Internacional de La Rioja), Ramón Salaverría (U. Navarra) y Astrid Wagner (CSIC), realizó un glosario de 125 términos sobre desinformación y odio que ayudarán a entender muchos aspectos.

El método DISARM se fundamenta en lo que se conoce como la pirámide de la desinformación (StratComCOE, & HybridCOE, 2021), siguiendo la línea de los estudios TTP. En la cúspide de esta pirámide se encuentra la ideación de las campañas, donde se define a los públicos objetivo. A partir de ahí, se crean los incidentes, se desarrollan las narrativas y, finalmente, en la base de la pirámide, se sueltan los artefactos de desinformación y/o odio. El defensor, es decir, quien busca desenmascarar estas campañas, debe seguir un proceso inverso. Primero,

debe identificar los artefactos de desinformación, luego analizar las narrativas, detectar los incidentes y, finalmente, identificar los objetivos primigenios de los ataques. Este enfoque es crucial para los que se dedican al *fact-checking*, es decir, a verificar la veracidad de mensajes en redes sociales y otros medios.

2.1. CÓMO SE COMPORTA LA OPINIÓN PÚBLICA EN REDES SOCIALES

Antes de adentrarnos en cómo se llevaría a cabo una campaña de desinformación y/o odio en internet hoy en día, es fundamental explicar ciertos aspectos que nos ayudarán a comprender cómo y en qué influyen las TTP que se abordarán más adelante. No se busca aquí elaborar un tratado sobre opinión pública y comportamiento psicológico, ya que existe una amplia bibliografía que trata estos temas en profundidad. En cambio, el objetivo es presentar de manera sencilla los factores que mayormente determinan el comportamiento en las redes sociales.

Lo primero que debemos entender es que no es necesario convencer a la totalidad de un grupo para modificar sus opiniones o lo que consideran aceptable. A pesar de la creencia generalizada de que se necesita convencer a una mayoría, la teoría sociológica de los enlaces débiles, propuesta por Granovetter (1973), nos ofrece una perspectiva diferente. Esta teoría sostiene que la difusión de información y la generación de oportunidades surgen más a partir de los enlaces débiles que de los fuertes. Los enlaces débiles se refieren a las conexiones y relaciones distantes y poco frecuentes que tenemos con conocidos o compañeros de trabajo, que nos permiten acceder a información y oportunidades nuevas y diversas. En contraste, los enlaces fuertes son aquellos que mantenemos con familiares y amigos íntimos, que aunque suelen ser muy solidarios, tienden a ser homogéneos en cuanto a la información y los recursos que aportan, lo que limita la diversidad de perspectivas. Así, resulta más fácil encontrar nuevas ideas al relacionarse con personas que vemos de vez en cuando o que conocemos por primera vez, en comparación con aquellas con las que interactuamos a diario. Esta dinámica también se aplica a la búsqueda de nuevas oportunidades, como conseguir un trabajo (es más probable que nos hable un amigo o un pariente lejano de la oportunidad de un nuevo trabajo que un amigo o familiar íntimo). Las conexiones ocasionales, de acuerdo a esta teoría, pueden abrir puertas a perspectivas y oportunidades que, de otro modo, podrían pasar desapercibidas en nuestras interacciones cotidianas.

En el contexto de las redes sociales, los usuarios se encuentran a menudo atrapados en lo que Eli Pariser denomina "filtros burbuja" (2012): en un mundo que se supone totalmente abierto, las personas tienden a buscar y escuchar sólo

aquello que coincide con sus ideas y refuerza sus opiniones, rechazando lo que las contradice o cuestiona. A ello hay que sumar que los algoritmos de las redes sociales e internet amplifican este fenómeno, lo que lleva a que cada vez más personas crean que tienen la razón y que su opinión es mayoritaria, aunque en realidad no lo sea, ya que solo interactúan con quienes piensan como ellos.

La búsqueda de opiniones y hechos que respalden nuestras propias creencias, mientras se evita la información que las contradice, se conoce como "sesgo de confirmación" en el ámbito de la psicología. Este fenómeno puede llevar a interpretaciones sesgadas de la realidad, ya que las personas tienden a interpretar la información en función de sus creencias preexistentes, lo que puede derivar en malentendidos, distorsiones o recuerdos selectivos. Por ejemplo, alguien puede ignorar opiniones políticas o medios de comunicación que presenten perspectivas opuestas, o seguir tratamientos alternativos o teorías conspirativas en lugar de aceptar la evidencia científica disponible sobre un hecho particular.

Figura 1: El sesgo de confirmación. Fuente: Paz (2020)

La teoría de los enlaces débiles ofrece un marco útil para explicar cómo se introducen nuevas ideas dentro de una burbuja de usuarios aparentemente homogénea. Según este enfoque, bastan unas pocas conexiones hacia el exterior para que un contenido distinto penetre en el grupo y empiece a circular, de modo que pequeños cambios en la estructura de la red pueden desencadenar transformaciones significativas en las opiniones compartidas (Myers, 2009). Aunque no existe un acuerdo cerrado sobre el umbral exacto necesario para que se produzca este vuelco, distintas investigaciones apuntan a que convencer a

alrededor de un 25-30% de los miembros puede bastar para que la mayoría termine adoptando esas nuevas ideas, siempre que existan las condiciones sociales y de red adecuadas. El mecanismo que explica esta dinámica combina la presión de conformidad con lo que en matemáticas y teoría de redes se denomina "ilusión de la mayoría". En determinadas configuraciones, cada individuo tiende a sobrestimar el apoyo que tiene una postura porque la observa repetida en su entorno inmediato, aunque en términos globales siga siendo minoritaria. Esa percepción distorsionada, unida al deseo de no desentonar con el grupo, lleva a muchas personas a alinearse con lo que creen que es la opinión dominante, fenómeno que se explota de forma sistemática en campañas políticas y estrategias de marketing comercial.

La Figura 2 ilustra este efecto con un ejemplo sencillo: se representan trece personas mediante círculos y las líneas indican quién se relaciona con quién (Grima, 2022). Once de esos nodos comparten una opinión (en blanco) y solo tres sostienen la opinión alternativa (en otro color), de modo que, a escala global, la mayoría es clara. Sin embargo, al observar cómo se distribuyen las conexiones, se comprueba que muchos de los nodos blancos tienen más vínculos con nodos de color que con otros blancos, de manera que, desde su punto de vista local, la postura minoritaria parece mucho más extendida de lo que realmente está. Aunque la mayoría objetiva sea blanca, una parte significativa de estos sujetos puede llegar a sentir que se encuentra en desventaja, desarrollando una visión errónea de la realidad que favorece la expansión del punto de vista minoritario.

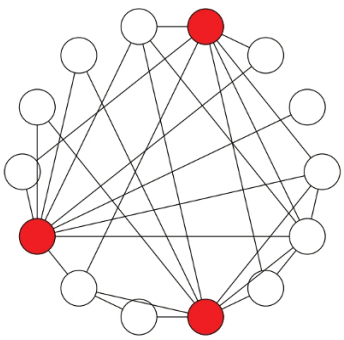

Figura 2: La ilusión de la mayoría. Fuente: Grima (2022)

Este tipo de configuraciones ha sido estudiado mediante modelos matemáticos y simulaciones en redes, con el objetivo tanto de amplificar como de mitigar la ilusión de mayoría según el contexto. Algunos trabajos se centran en identificar qué estructuras de vínculos débiles facilitan que una innovación o

una narrativa marginal se propague rápidamente, mientras que otros exploran cómo rediseñar las conexiones o introducir información correctiva para reducir la percepción distorsionada de apoyo. Estos resultados resultan especialmente relevantes para comprender por qué ciertas campañas de desinformación o de odio consiguen ganar tracción con recursos limitados, y qué tipo de intervenciones podrían ayudar a neutralizar su efecto antes de que logren consolidarse en el imaginario colectivo (Dippel, & et al., 2024).

Quienes discrepan de las ideas dominantes de su grupo pueden quedar atrapados en lo que se ha descrito como una "espiral del silencio", concepto formulado por la socióloga alemana Elisabeth Noelle-Neumann en la década de 1970 (Naelle-Nemann, 1995). Esta teoría sostiene que la percepción de cuál es la opinión mayoritaria (aunque a menudo se trate solo de una impresión difusa, no de un dato contrastado) condiciona fuertemente la disposición de las personas a expresarse en público, especialmente cuando se trata de temas controvertidos. Cuando alguien llega a la conclusión de que sus puntos de vista son minoritarios o socialmente poco aceptables, tiende a callarse para evitar el aislamiento, el rechazo o el descrédito, lo que reduce aún más la visibilidad de esa postura en el espacio público. A medida que más individuos silencian sus discrepancias, la opinión percibida como mayoritaria gana presencia y legitimidad, lo que refuerza la impresión de consenso y empuja a otros a adaptarse a esa corriente aparente, alimentando así un proceso circular en el que las voces disidentes se vuelven cada vez más raras y difíciles de sostener.

Asimismo, dentro de los grupos con los que nos relacionamos, ya sea en la vida real o en el ámbito digital, hay varios aspectos más que pueden influir significativamente en la opinión pública. A continuación, citamos algunos de estos factores, muchos de los cuales han sido expuestos por la Agencia de Defensa Psicológica del Gobierno de Suecia (Psychological Defense Agency, 2025) como trampas de pensamiento, junto con otros elementos adicionales que se aportan:

- La teoría de la persuasión: el cómo se comunican las ideas puede cambiar las opiniones. Un mensaje bien estructurado, bien presentado y con componentes lógicos y/o emocionales puede persuadir mucho. Así, el marketing político establece discurso puramente emocional para oyentes de bajo nivel educativo, mientras que mezcla emocional con un poco de lógica para públicos con mayor nivel (León, 1993). Un emisor con credibilidad, autoridad y confiabilidad, atractivo, con carisma y supuesta experiencia y conocimiento será más atendido, mientras que los mensajes deberán adecuarse a las características de los receptores

(edad, género, educación, creencias, opiniones, motivaciones y necesidades) así como del contexto (cultura y normas sociales del lugar).

- La teoría de la identidad social: desarrollada en los 70 (Tajfel, Billig, & Bundy, 1971), explica cómo las personas se definen a sí mismas en función de su pertenencia a grupos sociales. Esta identificación les ayuda a percibir su lugar en el mundo y a entender cómo interactuar y compararse con los demás. La pertenencia a un grupo fomenta la cohesión y el apoyo entre sus miembros, creando un sentido de unidad entre aquellos que comparten la misma identidad y creencias. Al mismo tiempo, esta identificación puede llevar a la discriminación hacia grupos externos, considerados diferentes. Es importante destacar que la identidad social no es estática, puede evolucionar con el tiempo a medida que las personas adoptan nuevas identidades y lealtades. Estas identidades pueden ser de naturaleza étnica, nacional, política o social, y su flexibilidad permite que las personas se adapten a diferentes contextos y experiencias a lo largo de sus vidas.
- Efecto de la polarización: se refiere a la tendencia de las personas a adoptar posturas más extremas que las que tenían inicialmente, especialmente en temas controvertidos como la política, la religión o cuestiones sociales. Esta dinámica se ve impulsada por la necesidad de alinearse con el grupo y de sentirse parte de él, lo que lleva a los individuos a adoptar opiniones más radicales, y factores como el sesgo de confirmación y la identidad social, que ya hemos discutido, juegan un papel crucial en este proceso amplificando la polarización. Existen varios elementos que contribuyen a esta tendencia, como son la presencia de discusiones (la presencia de continuas discusiones intensas, especialmente aquellas que atacan creencias e identidades profundamente arraigadas, pueden llevar a que las personas se aferren aún más a sus posturas), exposición selectiva (la falta de contacto con perspectivas externas, dentro de una "burbuja" informativa, limita la diversidad de opiniones y refuerza las creencias existentes), y retroalimentación (compartir y recibir opiniones extremas dentro de un grupo puede intensificar y reforzar las creencias, creando un ciclo de retroalimentación que hace que las posturas se vuelvan aún más radicales).
- Sesgo de disponibilidad: las personas sobreestiman la probabilidad y la importancia de aquello que recuerdan fácilmente en su memoria. Por eso si se sitúa algo de manera frecuente en las redes o medios y lo vemos a menudo, es más probable que creamos que es más común o relevante de lo que puede ser.

- Falso consenso: sobreestimar la cantidad de personas que comparten nuestras opiniones, valores o comportamientos. Las personas quieren creer que son la opinión mayoritaria.
- El efecto halo: se atribuyen cualidades a las personas si es buena en otra cosa. Por ejemplo, a una persona muy inteligente se le presupondrá que es además buena persona, o un *influencer* con aspecto un físico impresionante y deportista se supondrá que además será experto en nutrición y bienestar.
- Los algoritmos de las redes sociales e internet: diseñados para "vendernos" mejor cualquier cosa, ya sean productos materiales o ideas. Para lograrlo, las empresas han desarrollado técnicas de perfilado y segmentación de usuarios, que se describirán con más detalle más adelante. Este proceso implica asignar a cada usuario a un grupo, clúster o "burbuja" previamente identificado y valorado. A través de esta clasificación, cada persona se categoriza según sus comportamientos, creencias y formas de informarse e interactuar, teniendo como resultado que los usuarios son expuestos a una "realidad" que refuerza y moldea sus gustos, opiniones y creencias. Esto contribuye al sesgo de confirmación, donde los individuos piensan que tienen razón porque perciben que "todo el mundo" opina y desea lo mismo que ellos. Este fenómeno puede intensificar la polarización, ya que al sentirse parte de su grupo, los usuarios se convierten a su vez en reemisores de contenido al sentirse más seguras al ser parte de una mayoría, siendo conocido este proceso en el ámbito de la comunicación como "cámara de eco" o "efecto carro de la compra" en el de la psicología, donde las ideas y opiniones se repiten y refuerzan, alimentando el proceso de polarización. La estructura de las redes sociales potencia este ciclo, utilizando algoritmos que favorecen la viralidad de los contenidos, independientemente de su veracidad.
- Las personas que se sumergen en una espiral donde solo escuchan las mismas ideas y se rodean de los mismos grupos entran en lo que se conoce como el efecto *rabbit hole*, término que evoca la imagen de Alicia cayendo por el agujero de la madriguera en "Alicia en el país de las maravillas", que describe cómo se pueden perder en una realidad paralela, alejándose de perspectivas diversas y críticas.
- El efecto *gaslighting*: forma de manipulación psicológica que distorsiona la realidad, la memoria y el juicio de las personas. A través de mensajes emocionales intensos y la continua distorsión de la realidad, se niegan

hechos, se minimizan experiencias y se reinterpretan eventos pasados. Esto lleva a las víctimas a cuestionar sus propios sentimientos, percepciones y opiniones, aislándolas de sus círculos y grupos de referencia. Como resultado, las personas que sufren *gaslighting* pueden perder confianza en sí mismas y en sus decisiones, experimentando un estrés emocional que las hace sentir atrapadas en una situación incomprensible. Este aislamiento crea desconfianza y puede llevar a las víctimas a buscar apoyo en nuevos grupos, donde intentan conectar y ser validadas, pero en esta búsqueda se les puede introducir nuevos valores e ideas que justifiquen mensajes de odio y distorsionen realidades, creando estigmas hacia otros. El término *gaslighting* proviene de la obra de teatro "*Gas Light*", escrita en 1938, que posteriormente fue adaptada al cine en 1940 en Inglaterra y en 1944 en Estados Unidos. En esta historia, un esposo manipula a su esposa para que crea que está perdiendo la cordura. A través de esta manipulación, él busca hacerla más dependiente, manipulable e influenciable, sembrando dudas sobre su propio juicio y fomentando que dependa más de las opiniones de los demás. La película de 1944, protagonizada por Ingrid Bergman, le valió el Óscar a la Mejor Actriz, mientras que Angela Lansbury recibió una nominación al Óscar a la Mejor Actriz de Reparto. La obra y sus adaptaciones han dejado una huella duradera en la cultura popular, convirtiéndose en un referente para describir situaciones de manipulación psicológica.

Todos los aspectos mencionados si son aprovechados por parte de alguien interesado conducen, como se puede observar, hacia la polarización y la pérdida de confianza en instituciones clave como la política, la democracia, la ciencia, los medios de comunicación, las fuerzas del orden o las ONG, entre otros. Esta falta de confianza no siempre se dirige contra individuos específicos, como meteorólogos o investigadores de vacunas, sino que a menudo busca cuestionar la autoridad y el poder de las instituciones en su conjunto. Los ataques a estas instituciones, junto con la difusión de narrativas alternativas y teorías conspirativas que apelan más a las emociones que a la razón, alimentan la desconfianza. Cuando se siembra la duda sobre todo, se produce una crisis de la verdad en la que las personas terminan siendo escépticas incluso ante los hechos más básicos.

En ese clima, incluso los hechos más sencillos pueden ponerse en cuestión, lo que desemboca en una auténtica crisis de la verdad, pues ya no se discuten interpretaciones o soluciones, sino la propia existencia de una realidad compartida sobre la que construir acuerdos. Este deterioro del consenso hace

mucho más difícil abordar los problemas colectivos de forma eficaz, y eso no es un efecto colateral, sino el objetivo de muchas de estas dinámicas: desestabilizar el sistema hasta que una parte significativa de la población llegue a la conclusión de que "nada funciona". Cuando se extiende la percepción de que las instituciones están irremediablemente rotas, las sociedades tienden a paralizarse o a aceptar soluciones de carácter extremo que se presentan como únicas salidas posibles. En ese contexto de cansancio y frustración, las respuestas simples a problemas complejos resultan especialmente seductoras, aunque a medio y largo plazo suelen traducirse en decisiones dañinas para la convivencia democrática y el bienestar colectivo.

2.2. LA PLANIFICACIÓN DE LA CAMPAÑA DE DESINFORMACIÓN Y ODIO

Dentro de las distintas fases, la inicial consiste en planificar qué se quiere conseguir, estableciendo una serie de objetivos a marcar y alcanzar al final del proceso. Con ello se pretende establecer las bases sobre las cuales se van a desarrollar las campañas a elaborar. Se deben establecer tres niveles en esta planificación:

1. Planificar los objetivos.
2. Planificar la estrategia a emplear.
3. Análisis de la audiencia objetivo.

2.2.1. Planificar los objetivos

Lo que se pretende es establecer de manera clara y medible objetivos que sean alcanzables. Es algo totalmente análogo a los objetivos dentro del marketing comercial, donde se siguen las reglas SMART, correspondientes a las siglas en inglés de específico *(Specific)*, medible *(Measurable)*, alcanzable *(Achievable)*, realista *(Realistic)* y limitado en el tiempo *(Time-bound)*. Así, el objetivo que se plantee tiene que ser algo realista, que pueda hacerse, que pueda medirse cómo evoluciona y si llega finalmente a los objetivos fijados, con un tiempo marcado desde el inicio para su realización. El enunciado no debe especificar ni la forma ni los medios para conseguirlo, y el efecto debe ser distinguible del objetivo.

Entre los objetivos suelen enmarcarse una serie de puntos claros en el ámbito que nos ocupa:

- Facilitar propaganda: Promover los intereses de un determinado grupo o estado.
- Degradar al adversario: Dañar la imagen y reputación de los oponentes.

- Contraatacar críticas: Responder a las críticas atacando a los actores que las emiten o acusándolos de tener opiniones sesgadas o extremistas.
- Desacreditar fuentes creíbles: Socavar la confianza en las instituciones, lo que facilita la influencia posterior.
- Distorsionar la realidad: Presentar realidades y verdades "alternativas" que confunden a la audiencia.
- Distraer: Desviar la atención hacia narrativas y personas diferentes.
- Amenazar a críticos: Intimidar a periodistas, académicos y otros para que dejen de informar.
- Dividir: Crear y/o ampliar conflictos entre grupos, aprovechando tensiones y hechos existentes.

Veámoslo con un ejemplo, imaginemos que el objetivo es reducir la credibilidad y reputación de un medio de comunicación (desacreditar fuentes creíbles) en un 20% en un plazo de 6 semanas. Para lograrlo, se debe establecer un objetivo SMART, que en este caso sería claro y conciso: reducir la credibilidad de ese medio de comunicación en un porcentaje determinado. Para medir el éxito, se podrían utilizar indicadores como el número de publicaciones en redes sociales que difunden información falsa y engañosa, el número de personas alcanzadas por la campaña y el cambio en la percepción pública del medio de comunicación objetivo a través de las quejas y el estado emocional de sus comentarios, medido a través de encuestas y análisis de sentimiento en redes sociales. Un objetivo así es cuantificable, lo que permite evaluar el progreso y ajustar la estrategia según sea necesario. También es realista, considerando que se puede lograr mediante la difusión de información falsa y engañosa en un plazo determinado. Además, es relevante para la campaña de desinformación, ya que busca influir en la percepción pública de un medio de comunicación, además de tener un plazo determinado, 6 semanas, lo que permite planificar y ejecutar la campaña de manera efectiva.

2.2.2. Planificar la estrategia a emplear

En este punto, se definen claramente los objetivos que se pretenden alcanzar, lo que implica determinar la audiencia objetivo. Este proceso, conocido en marketing como segmentación, consiste en clasificar a la población en distintos grupos y establecer cuál será el de interés al que se dirigirá principalmente la campaña. Es importante tener en cuenta que no toda la población responderá

de la misma manera ante los mismos estímulos, pues lo que puede resultar efectivo para un grupo puede no funcionar con otro. Por lo tanto, la segmentación permite adaptar los mensajes y estrategias a las características específicas de cada grupo, maximizando así la efectividad de la campaña.

La clasificación de personas o grupos se puede realizar históricamente de dos maneras: mediante el conocimiento previo que se tiene de la población o posterior a partir de datos existentes. En la actualidad, debido a la gran cantidad de datos disponibles que ofrece internet y especialmente las redes sociales y teléfonos móviles, la clasificación previa ha quedado en gran medida obsoleta. Ahora, se basa principalmente en los datos extraídos a partir de una serie de parámetros de perfilado de los usuarios en redes sociales, proceso de clasificación que puede llevar tiempo, ya que implica el seguimiento de cuentas y la recopilación de información. Sin embargo, los tiempos pueden acortarse, ya que existen mercados donde se comercializan los datos y características de numerosas cuentas de usuarios en redes sociales. Estos datos a menudo se venden en la *dark web*, lo que permite a las empresas y otros interesados acceder a información valiosa para sus campañas.

No todos los grupos dedicados a la segmentación cuentan con las mismas capacidades de recolección y cantidad de datos. Tras el Brexit y las elecciones de Estados Unidos en 2016, las redes sociales han limitado o dificultado el acceso a estos datos, lo que ha llevado a que se necesiten recursos mucho más grandes y potentes para obtenerlos.

Una de las formas más comunes de recolección de datos es a través del uso de miles de cuentas que se hacen pasar por usuarios reales en redes sociales y foros de diarios digitales, participando en diversos temas, generalmente de manera bastante emocional y activa. En investigaciones propias sobre cuentas identificadas como grandes propagadoras de odio en la red X/Twitter, se observó que estas cuentas solo compartían contenido de odio o campañas desinformativas siempre entre el 10% y el 12% de sus publicaciones, ni más ni menos. Estas cuentas suelen ser lo que denominamos *nanoinfluencers*, es decir, tienen entre 100 y 1.000 seguidores, teniendo además fotografías que suelen ser genéricas o relacionadas con los temas que abordan. Casi el 90% de sus mensajes restantes se centran en temas casi monotemáticos, como gatitos y otros animales, fútbol, fiestas de la ciudad, música, televisión, fuerzas del orden, entre otros. Esto hace que otros usuarios las perciban, por ejemplo, como seguidores de su equipo de fútbol, lo que puede llevar a que les den un "me gusta" o "*like*".

A partir de este punto, estas cuentas buscan establecer un seguimiento mutuo, lo que les permite acceder a una mayor cantidad de datos de las cuentas

interconectadas. Una estrategia interesante que emplean para fomentar estos contactos es generar discusiones entre grupos que aparentan ser opuestos, utilizando lo que se conoce como operaciones de falsa bandera. En este escenario, grupos que se hacen pasar por el contrario crean una confrontación aparente. Esto lleva a que los usuarios que interactúan con cualquiera de los bandos en disputa sean categorizados, siguiendo a aquellos que muestran "me gusta" en uno u otro lado. Una vez que han clasificado y recopilado datos de las cuentas que han interactuado con ellas, estas cuentas suelen desprenderse de algunos seguidores. Esto lo hacen para no parecer que tienen un número excesivo de personas siguiendo a sus seguidores y seguidos, manteniéndose así en el rango de los *nanoinfluencers*. Sin embargo, en tiempos recientes, se ha observado que estas cuentas comienzan a aumentar el número de seguidores y seguidos que aceptan. Esto se debe a que muchas de ellas llevan más tiempo en la red y, por lo tanto, generan una mayor credibilidad.

Una vez que han establecido contacto a través de una cuenta, comienzan a recopilar una variedad de datos sobre ella. Esto incluye desde los mensajes y "me gusta" hasta los enlaces, información y fotografías que el usuario comparte en sus redes sociales. Sin embargo, el análisis de fotografías y vídeos requiere equipos e infraestructuras muy avanzadas, que generalmente solo están al alcance de grandes organizaciones con objetivos específicos. Por ejemplo, pueden enfocarse en usuarios que critican un régimen totalitario en su país. Por el momento, el análisis de imágenes y vídeos se limita generalmente a aspectos más básicos, como el estudio de los colores utilizados. Esta información puede ofrecer pistas valiosas sobre el perfil del usuario detrás de la cuenta, ayudando a construir un retrato más completo de sus intereses y comportamientos.

Las segmentaciones de perfilado de los usuarios se llevan a cabo mediante técnicas informáticas de aprendizaje automático. A partir de los datos recopilados, se aplican algoritmos que permiten clasificar y agrupar la información de manera matemática, donde el algoritmo comienza con una variable objetivo (o dependiente) que se desea analizar y, a partir de ahí, establece la importancia y la relación de las demás variables (independientes) y datos encontrados para determinarla. Este proceso permite clasificar de manera precisa grupos de comportamientos, reacciones, gustos y opiniones similares. Así, se pueden identificar patrones que ayudan a comprender mejor a los usuarios y a personalizar la comunicación de forma más efectiva.

De este modo al realizar, por ejemplo, una encuesta, ya no se limita a obtener resultados generales como "los hombres opinan esto y las mujeres aquello" o "los mayores frente a los jóvenes" de manera bastante global. Ahora se puede

ser mucho más específico. Por ejemplo, se puede afirmar que los jóvenes de un determinado nivel educativo y que residen en áreas con un nivel adquisitivo específico tienen respuestas diferentes en comparación con aquellos de otro barrio que presentan un menor nivel educativo y adquisitivo e ideológico. Esta capacidad de segmentación permite un ajuste mucho más fino y preciso. Cuantos más datos se recopilen y empleen, más exactas serán las conclusiones que se puedan extraer, lo que facilita una comprensión más profunda de las diferencias en opiniones y comportamientos entre distintos grupos.

Un ejemplo de venta de datos segmentados era el de la empresa Insights+ en Estados Unidos, que hasta 2023 ofrecía datos detallados sobre los habitantes de una zona extraídos principalmente a partir de redes sociales, así como por geolocalización de sus teléfonos móviles. Esta información permitía a predicadores religiosos conocer aspectos como las posturas espirituales de la comunidad, si estaban casados, si utilizaban la Biblia, en qué áreas residían e incluso si tenían alguna posible adicción. La capacidad de Insights+ para proporcionar este tipo de datos ilustraba cómo la segmentación puede ser utilizada para adaptar mensajes y estrategias a las características específicas de un público objetivo.

2.2.3. Análisis de la audiencia objetivo

Una vez que se han recogido y segmentado los datos de la población, se procede a estudiar las características de los grupos identificados. Existen varias variables que resultan especialmente útiles en las campañas, y una de las más importantes es la segmentación geográfica.

Los datos que los usuarios comparten en sus redes sociales y comentarios permiten incidir en áreas muy específicas, lo que resulta valioso. Las personas tienden a establecerse en zonas de ciudades y pueblos que reflejan su situación económica, social y educativa. Por ejemplo, una población en barrios económicamente desfavorecidos, donde predominan trabajos de mano de obra y hay una fuerte presencia de inmigrantes, puede convertirse en un foco particular para campañas que fomenten el odio hacia estos grupos. Esta segmentación geográfica no solo ayuda a identificar el contexto de la población, sino que también permite diseñar mensajes que resuenen con las realidades y preocupaciones de cada comunidad.

Acompañada a la segmentación geográfica se añade la segmentación demográfica, en la que la edad, el género y su disponibilidad económica van a ser igualmente puntos importantes. No es lo mismo acceder a jóvenes adolescentes, que se mueven en redes sociales como Instagram o TikTok, con un mensaje

corto y más visual y unos intereses movidos por el empleo, las ganancias de dinero o la vivienda, a gente más mayor que se mueve principalmente en Facebook donde los textos pueden ser un poco más elaborados y se centran en bienestar, aspectos sociales más generales y la salud. La determinación del sexo de la persona de una determinada cuenta es fácilmente calculable a partir de los elementos que comparte y permite igualmente ajustar posteriores mensajes y clasificaciones.

Una de las segmentaciones más avanzadas, y a la par más controvertidas y expresamente prohibidas por la legislación europea, es la psicográfica. Es importante recordar que la empresa Cambridge Analytica, la que trabajó en el Brexit y la primera campaña de Trump en Estados Unidos, utilizando tan solo 68 "me gusta", era capaz de predecir características como el color de piel con un 95% de fiabilidad, la orientación sexual con un 88% y las inclinaciones políticas con un 85%. Además, podían estimar el nivel de inteligencia, la religión, el uso de tabaco y otras drogas, e incluso deducir si una persona provenía de un hogar con padres divorciados. El método más utilizado por Cambridge Analytica fue el modelo OCEAN, que se basa en cinco grandes dimensiones de la personalidad: apertura, responsabilidad, extraversión, amabilidad y neuroticismo. Según los psicólogos de las Universidades de Cambridge y Stanford que colaboraron en los estudios iniciales con OCEAN, este enfoque permite categorizar la personalidad de una persona a partir de sus "me gusta" en Facebook con mayor rapidez y precisión que un psicólogo utilizando cuestionarios tradicionales, superando su efectividad en un 7% (Park, & et al., 2015). Incluso las preferencias musicales son utilizadas en la medida de estas variables de la personalidad, ya sea a través de los me gusta en artistas (Nave, & et al., 2018) o sobre listas que se escuchan en portales de música. Pueden probar ustedes mismos con los datos de sus redes sociales en la web de *The Psychometrics Centre* de la Universidad de Cambridge para obtener un perfilado OCEAN según lo que publican en Twitter, Facebook y Linkedin, o incluso un texto abierto (https://applymagicsauce.com/demo).

Estos modelos han sido probados en diversos países, como Italia, España, Alemania, Grecia y Polonia, para verificar su efectividad en diferentes sociedades. Los resultados indicaron que era más fácil predecir el comportamiento de una persona a partir de su personalidad deducida que únicamente a partir de variables demográficas como la edad, el sexo, los ingresos o el nivel de estudios. Pero la predicción se vuelve aún más precisa cuando se combinan todas estas variables en conjunto. Curiosamente, se encontró que la apertura era el mayor predictor de la ideología de una persona (Vecchione, & et al., 2011), e igualmente, se investigó qué tipos de

personalidad serían más susceptibles a la desinformación, indicando los resultados que las personas más extrovertidas, así como aquellas que son menos concienzudas y agradables, tienden a ser más propensas a creer y compartir información errónea. También se observó que los rasgos de la "tríada oscura" (narcisismo, psicopatía y maquiavelismo) están positivamente asociados con la difusión de información errónea. En particular, el narcisismo y la psicopatía se relacionan con una mayor creencia en la desinformación (Calvillo, León, & Rutchick, 2024). Además del modelo OCEAN, algunos investigadores también trabajan con otros enfoques teóricos de personalidad, como el MBTI, que se basa en cuatro variables: extraversión/introversión, sensación/intuición, pensamiento/sentimiento y juicio/percepción (Stajner, & Yenikent, 2021), aunque es mucho menos usado.

El perfilado psicográfico o psicológico de las cuentas permite crear mensajes mucho más personalizados que cualquier otro tipo de clasificación. Cambridge Analytica llegó a identificar más de 40 tipos de personalidad, pero para ilustrar su enfoque, consideremos 5 tipos de personalidad y cómo una inteligencia artificial podría generar un mensaje persuasivo que convenza a alguien de que un medio de comunicación X no es creíble. Partimos de 5 tipos de personalidad construidos a partir de los datos de OCEAN:

- El escéptico: este tipo de persona tiende a cuestionar la información que recibe. Un mensaje dirigido a ellos podría resaltar inconsistencias en las noticias del medio, invitándolos a investigar más a fondo.
- El conformista: aquellos que valoran la aceptación social. Un mensaje para este grupo podría enfatizar que muchos de sus amigos o conocidos desconfían de ese medio, sugiriendo que seguir esa opinión es lo más sensato.
- El aventurero: personas que buscan nuevas experiencias y emociones. Un mensaje para ellos podría apelar a su deseo de descubrir la verdad, sugiriendo que el medio en cuestión oculta información importante.
- El analítico: este tipo de personalidad se basa en datos y hechos. Un mensaje dirigido a ellos podría incluir estadísticas y estudios que demuestren la falta de credibilidad del medio.
- El empático: aquellos que se preocupan por los demás y las injusticias. Un mensaje para este grupo podría centrarse en cómo el medio ha fallado en representar adecuadamente las voces de las comunidades marginadas.

Los mensajes que nos propone una IA para cada uno son:

- El escéptico: "¿Te has dado cuenta de las contradicciones en las noticias de X? A menudo presentan información que no se sostiene al contrastarla con otras fuentes. Te animo a que investigues más y compares lo que dicen con datos de medios más confiables. La verdad siempre merece ser buscada."
- El conformista: "Muchos de tus amigos y conocidos han comenzado a cuestionar la credibilidad de X. La mayoría de las opiniones en tu círculo sugieren que este medio no siempre presenta la verdad. ¿Por qué no unirte a ellos y considerar otras fuentes que son más respetadas y confiables?"
- El aventurero: "¿Te gustaría descubrir la verdad detrás de las noticias? El medio X ha sido criticado por ocultar información crucial y presentar una narrativa sesgada. Te invito a explorar otras fuentes que te ofrezcan una visión más completa y emocionante de la realidad."
- El analítico: "Los datos no mienten. Estudios recientes han demostrado que el medio X tiene un historial de reportar información inexacta y sesgada. Al analizar sus informes, verás que muchas afirmaciones carecen de respaldo. Te animo a que revises las estadísticas y compares con fuentes más objetivas."
- El empático: "Es preocupante cómo el medio X ha fallado en representar adecuadamente las voces de las comunidades más vulnerables. Al ignorar sus historias y realidades, este medio no solo pierde credibilidad, sino que también perpetúa injusticias. Es fundamental buscar fuentes que realmente se preocupen por la verdad y la equidad."

Hay más ejemplos, académicos de diversas universidades de EEUU, Suiza e Inglaterra diseñaron y publicaron en 2019 un modelo algorítmico para determinar, a partir de los datos que los usuarios publican en Facebook, predecir sus ingresos para usos económicos y sociológicos (Matz, & et al., 2019).

Otro tipo de segmentación se basa en los medios de comunicación y las noticias que una cuenta sigue y comparte. Es importante tener en cuenta que cada medio tiene una línea editorial ideológica, lo que crea una burbuja de comunicación con seguidores que comparten ciertas creencias y valores. Si varias cuentas difunden enlaces de determinados medios y estas cuentas pertenecen, según el algoritmo clasificador, a un grupo o clúster específico, se puede inferir que todo el grupo tiende a tener una misma inclinación ideológica. Además, es fundamental

recordar que la tendencia ideológica política, junto con factores como la personalidad, la geografía y la demografía, permite un perfilado muy profundo en la clasificación de cada usuario. Esta combinación de datos ayuda a entender mejor las motivaciones y comportamientos de las personas en el entorno digital.

Por último, existe la segmentación basada en la estructura y el comportamiento de una cuenta dentro de su red de contactos y relaciones en las redes sociales. Para ello, se aplica lo que se conoce como la teoría de redes o de grafos, una rama de las matemáticas que se originó a partir del trabajo del matemático Euler. Este fue desafiado a encontrar una manera de cruzar todos los barrios de la ciudad prusiana de Königsber (actualmente Kaliningrado, Rusia) sin pasar dos veces por ninguno de los siete puentes de la ciudad, y para demostrar que era imposible inició el desarrollo de un nuevo campo de las matemáticas.

La teoría de redes estudia cómo se relacionan los datos que presentan conexiones entre sí. En el ámbito de la comunicación y las redes sociales, estos estudios permiten identificar grupos (clústeres) y también a los nodos (usuarios) que emiten más mensajes (conexiones o aristas). Se pueden identificar los usuarios más importantes e influyentes en la estructura del grupo (vector propio o *eigenvector*), y aquellos que actúan como puentes entre diferentes grupos, que aunque puedan participar poco, su ausencia podría dificultar la difusión de un mensaje entre distintos sectores (intermediación o *betweenness*).

Al analizar campañas anteriores utilizando la teoría de redes, es posible determinar matemáticamente la influencia de una cuenta en su entorno y grupo. Si se convence o se ataca a una de estas cuentas, su impacto será más profundo y rápido que si se actúa sobre otras cuentas menos conectadas. Las cámaras de eco que siguen a esas cuentas influyentes se verán rápidamente afectadas por lo que pase.

Como ejemplo, en la Figura 3 se puede observar la difusión de 119.640 mensajes provenientes de 31.258 cuentas en Twitter del diario digital Periodistadigital.com y estudiado en un artículo académico junto a la difusión de la web Caso Aislado, portales conocidos por su difusión de campañas de desinformación en España y promoción de ideas de extrema derecha (Arce García, Vila Márquez, & Fondevila i Gascón, 2021). En este gráfico (o mejor dicho grafo), cada punto representa una cuenta, y las líneas que las conectan indican un retweet o reenvío de un mensaje. El algoritmo agrupa por colores las cuentas con comportamientos similares, destacando en letras más grandes las cuentas con mayor vector propio o importancia en la red. Se puede notar que, además de la cuenta oficial del diario, hay otras cuentas clave en la difusión sobre las que vertebran mensajes y comparten apoyos.

También se observa que existen cuentas que redifunden mensajes de manera compartida y simétrica, lo que sugiere un grupo muy egocéntrico que rota en el reenvío de mensajes de las cuentas principales. En naranja, a la derecha del grafo, se puede ver una cuenta opositora en su discurso junto con sus seguidores. Un análisis del historial de creación de estas cuentas participantes revela que la mayoría se establecieron en 2011, en octubre de 2017 (coincidiendo con el proceso catalán) y a comienzos de 2020 (al inicio de la pandemia por Covid-19).

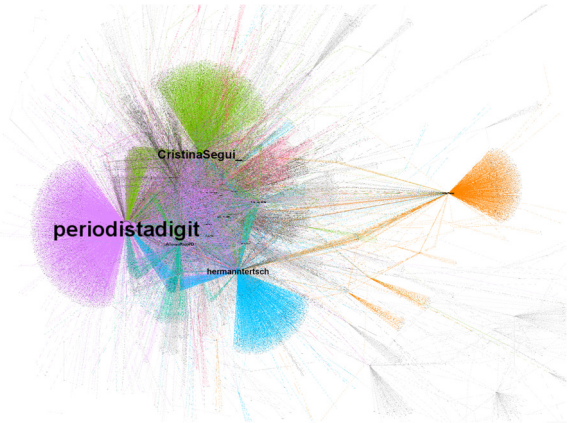

Figura 3. Difusión de retweets en Twitter del portal PeriodistaDigital.com en el año 2020.

La identificación de segmentos de estas cuentas y grupos permite, además, "jugar" con ellos en lo que se conoce como el "cultivo de activos ignorantes" de determinados segmentos (en inglés se suele emplear la expresión más despectiva *useful idiots*). Un ejemplo claro de esta estrategia se observa en el caso de Curro (Vila Márquez, & Arce García, 2019), un perro de pura raza alana española que iba a ser sacrificado a finales de 2018 y que se convirtió en el principal *trending topic* en España durante la Navidad de ese año. Según los portales de noticias Casoaislado.es y Despierta.info, se afirmaba que Curro había defendido su hogar de un ladrón rumano con un historial delictivo de 70 detecciones, y que iba a ser sacrificado por haberle arrancado varios dedos al intruso. Sin embargo, ni el perro ni el ladrón existieron realmente, pero la campaña resultó ser muy efectiva. Lo interesante de este caso es la identificación del segmento animalista, especialmente en torno al partido PACMA, como un elemento difusor de la información de manera ignorante. En las primeras fases de la campaña, se logró que muchos seguidores animalistas amplificaran la historia de forma activa, incluso con recogida de firmas en la plataforma Change.org, sin darse cuenta de que estaban siendo utilizados para propagar un

mensaje falso. Cabe destacar que los primeros mensajes que hablaron de Curro y su "sacrificio", fueron cuentas que informaban sobre el tráfico de Caracas, Venezuela.

Figura 4. Portadas del día 21 de diciembre de 2018 en despierta.info y casoaislado.es

Como se puede comprobar, los datos que proporcionan las redes sociales, internet y los dispositivos móviles permiten categorizar a los usuarios con un nivel de detalle sin precedentes. Gracias al aprendizaje automático y la inteligencia artificial, es posible segmentar a niveles muy detallados y crear mensajes personalizados para cada individuo de manera sencilla y rápida. Más aún si se dispone de equipos con inteligencia artificial incorporada que automaticen todos estos procesos de identificación, segmentación y respuesta, representando una herramienta de muchísimo valor para los organizadores de campañas: les permite conocer a fondo a sus objetivos, ya sea para incentivarlos o desincentivarlos en alguna acción, y adaptar el mensaje específicamente a cada uno aprovechando y explotando las ideas y prejuicios que contengan.

La segmentación permitirá escoger cuál será el público objetivo, cuál evitar, cuál utilizar, y cada uno tendrá sus respectivas ideas, cultura, costumbres o religión, pero también sus prejuicios, fisuras, ideas conspiranoicas, cámaras de eco o seguidores, vulnerabilidades, adversarios, amigos, redes sociales que emplean, diarios que leen, etc.

Esta segmentación tan precisa y el análisis matemático de las conexiones entre cuentas permiten también estudiar cómo se articulan las campañas de desinformación en red. Otro ejemplo es el canal de Telegram de Elecciones Transparentes, una asociación que sostiene que las elecciones en España están siendo manipuladas y que reproduce marcos discursivos ya vistos en otros países

(Rodríguez-Fernández, González-Fernández, & Arce-García, 2025). Si se siguen en un par de pasos los enlaces que difunde este canal, aparece una constelación de cuentas donde, aunque exista un nodo muy emisor, la verdadera columna vertebral de la red de enlaces se organiza en torno a otros canales.

Los nodos con mayor centralidad (*eigenvector* o vector propio) que destacan en la red de enlaces del canal, por este orden, son los del canal "Alvise Pérez", eurodiputado condenado en varias ocasiones por difundir insinuaciones sin aportar pruebas y cuya trayectoria pública se ha entrelazado inicialmente con la de Toni Cantó (Ballesteros, 2020; Pozas, 2023). Le siguen el canal "Noticias Rafapal", un canal conocido por la difusión sistemática de contenidos conspirativos, antisistema y pro-Kremlin (Olmo, 2023); "LIBERUMASOCIACION", una asociación que se presenta como defensora de los derechos y libertades "usurpados" durante la pandemia y que ha evolucionado hacia un discurso contra el "Estado abusador" (Algaba, 2025); y el canal "Julio Ariza", que por el nombre correspondería al fundador del grupo de comunicación Intereconomía y figura mediática (González, 2023). El mapa de conexiones muestra así cómo un canal dedicado a cuestionar la limpieza de los procesos electorales termina enlazado con un ecosistema más amplio de actores mediáticos y políticos que comparten marcos comunes. Este grupo de "Elecciones Transparentes" igualmente expone vídeos en Youtube, Rumble y Odysee, estas dos últimas plataformas de vídeos para figuras y contenidos que generalmente han sido prohibidos en otras plataformas. Rumble estaría aliado desde finales de 2021 al grupo *Trump Media & Technology Group*, y tendría entre sus inversores a J.D. Vance y Peter Thiel (Hagey, 2021; Siegelman, 2022), mientras que Odysee se asocia a vídeos de desinformación y odio, y es por donde emiten los canales rusos RT y Sputnik tras su prohibición en Europa tras la guerra de Ucrania (Martin, 2022; Marshall, & Tanfani, 2022).

2.3. LA PREPARACIÓN DE LA CAMPAÑA

Una vez terminado el plan, en el que se establecieron los objetivos, identificó, segmentó y analizaron los distintos grupos en la opinión pública que van a ser motivo de la campaña, se pasa a su preparación. Según el esquema DISARM que estamos siguiendo como guía, tendríamos los siguientes puntos:

- Desarrollar las narrativas: se establecen cuáles van a ser las narrativas más eficaces que alcancen mayor cobertura social, tanto *online* como *offline*.
- Desarrollar los contenidos: creación de los textos, imágenes, memes, y/o vídeos que van a emplearse en la campaña, así como establecimiento de los *hashtags* y palabras clave.

- Establecer activos sociales: se deben de crear páginas webs, diarios digitales, podcast, uso de influencers, estructuras de redes ya introducidas anteriormente, infiltraciones en otras burbujas o grupos, cultivo de activos ignorantes, recaudador de fondos, etc.
- Establecer legitimidad: creación de falsos expertos, justificaciones académicas o pseudocientíficas, creación de sitios de noticias falsos, suplantación de identidades, generación de campañas tipo *astroturfing*, cuentas parodia o engaño de fuentes confiables.
- *Microtarget*: establecimiento de objetivos muy específicos, muy dirigidos y localizados.
- Seleccionar canales y posibilidades: establecer cuáles son los lugares por los que la campaña se va a desarrollar y medios de apoyo (encuestas, audios, chats, videos, retransmisiones, foros, redes sociales, blogs, medios de comunicación tradicionales, canales diplomáticos).

2.3.1. Desarrollo de narrativas

Los mensajes a lanzar no pueden ser una sucesión inconexa de textos que busquen impactar sin más. Una buena campaña tiene que contar una historia para tener éxito, y será la que verdaderamente conecte con el público objetivo. Esta manera de contar una historia, que ofrece mensajes dentro de un contexto y establece unas narrativas, se denomina el *storytelling* en el ámbito de la comunicación. Y a la hora de contar historias existe un modelo muy empleado en la literatura, pero también en el ámbito de los ingenieros sociales de la desinformación y el odio: el esquema de Freytag (Barbu, & et al., 2025). En las redes no hay una novela unitaria, pero sí micro y macro-historias que pueden leerse como narrativas distribuidas, donde cada mensaje, artículo o hilo de tuits aportan fragmentos de la narrativa (Elkins, 2025). El esquema de Freytag sigue los siguientes puntos, separados en distintos actos que pueden acortarse o alargarse en el tiempo según los sentimientos y picos de atención observados en la población:

1) Exposición (EXP): se presenta el contexto básico de la historia (quién, dónde, cuándo) y la situación inicial antes del conflicto. Se identifican víctimas, héroes y villanos, así como se normalizan contextos.
2) Acción ascendente (RA): aparecen incidentes desencadenantes, conflictos y complicaciones que aumentan la tensión y llevan hacia un punto

álgido. Son mensajes que añaden acusaciones, incidentes, nuevas evidencias, dramatizan riesgos y llaman a la acción, mediante micromensajes acumulativos.

3) Clímax (CL): momento de máxima tensión o giro, donde se decide el rumbo del conflicto principal, como un anuncio oficial, filtración clave, ataque, decisión judicial, etc. Se produce el pico emocional de miedo, ira o indignación.
4) Acción descendente (FA): se muestran las consecuencias directas del clímax y empieza a "desenredarse" la trama. Se producen aclaraciones, rectificaciones, contramedidas, declaraciones que bajan la tensión. Se especifica y refuerza el quiénes somos, cómo funciona el mundo y qué deberíamos hacer, se afianzan confirmaciones sesgadas.
5) Desenlace (DE): cierre de la historia, resolución del conflicto y retorno a una situación relativamente estable, con informes finales, acuerdos, olvido mediático y cambio de agenda o temas a tratar. Se consolida una nueva realidad como verdad.

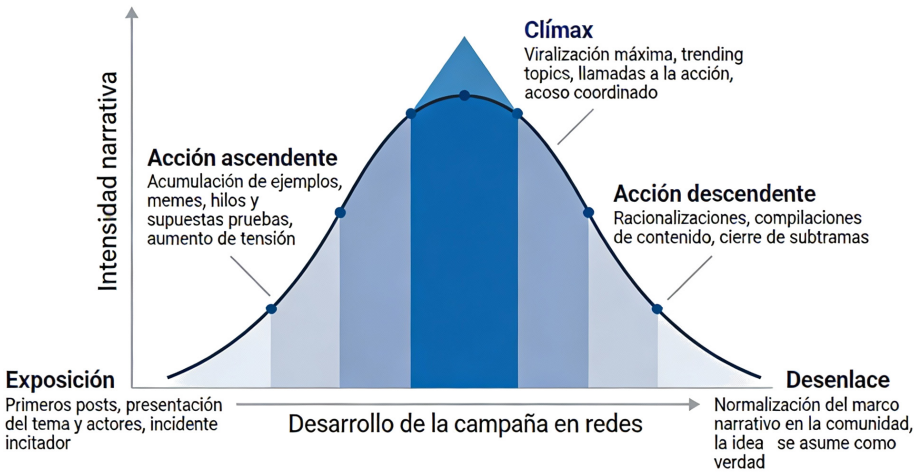

Figura 5. Esquema o pirámide de Freytag aplicada a campañas de desinformación y discurso de odio en redes sociales. Fuente: Elaboración propia.

A la hora de concretar cuáles son los textos, imágenes o vídeos que se van a ofrecer, se deben de tener en cuenta una serie de variables que van a favorecer el impacto de la misma entre la opinión pública objetivo. Todas y cada una de ellas servirán para

crear mayor impacto emocional, aprovechando las vulnerabilidades, costumbres y creencias de dicha sociedad. Veamos las principales (La Moncloa, 2025):

- Aprovechar las vulnerabilidades de la sociedad: la existencia de minorías, divisiones internas y externas, un ecosistema mediático frágil y unas instituciones públicas que han perdido credibilidad son elementos que se explotan con frecuencia (Jeangène Vilmer, & et al., 2018). Existen numerosos ejemplos de esto, como el aprovechamiento de poblaciones de lenguas minoritarias, el uso de la población inmigrante (especialmente si proviene de culturas y religiones muy diferentes), y la exposición de corrupción. Además, los medios de comunicación polarizados y los tabloides amarillistas que se presentan como medios tradicionales, junto con instituciones públicas ineficaces, se convierten en chivos expiatorios para redirigir la atención hacia ideas que acentúan las divisiones y la polaridad en la sociedad. Este enfoque no solo alimenta la desconfianza, sino que también profundiza las fracturas sociales, dificultando la cohesión y el entendimiento entre diferentes grupos.

Figura 6. Campañas de los partidos políticos Chega (Portugal) y Vox (España) en Instagram el mismo día 03/06/2024 para las elecciones europeas (Casquinho, & et al., 2024).

- Narrativas competitivas: discursos que niegan, relativizan o reinterpretan los hechos, generando versiones paralelas o "realidades alternativas" que desorientan al público. Su propósito es provocar conversaciones contradictorias y profundizar la polarización, sembrando la duda sobre lo que es cierto a través de una avalancha de argumentos, exigencias de prueba y flujos de información errónea. Este mecanismo, conocido como "manguera de falsedades" o *firehosing*, inunda de forma deliberada el espacio público con datos

distorsionados hasta volver casi imposible distinguir entre lo verídico y lo falso. Más allá de la confusión inmediata que produce, esta estrategia mina la confianza en las fuentes de información y debilita la capacidad colectiva de construir criterios compartidos o juicios bien fundamentados.

- Narrativas conspirativas: aprovechando la necesidad humana de comprender el mundo a través de sus propias creencias y conocimientos, así como la tendencia a simplificar lo que sucede, surgen narrativas que invocan la existencia de actores poderosos, seres o eventos que introducen confusión y desafían el orden establecido. En un contexto donde la opinión pública está poco informada, marginada o inclinada hacia perspectivas alternativas, se crean o amplifican hechos conspirativos que actúan como un complemento perfecto para las narrativas competitivas. Estas teorías no solo ofrecen explicaciones simplistas a situaciones complejas, sino que también alimentan la desconfianza hacia las instituciones, científicos, académicos y los medios de comunicación, reforzando así la percepción de que hay fuerzas ocultas en juego.

Figura 7: Narrativa conspiranoica de *chemtrails* (supuesta fumigación del cielo que altera el clima).

- Respuesta ante noticias de actualidad y crisis: los momentos iniciales de una crisis o de una noticia de última hora son oportunidades clave para influir en la opinión pública, que a menudo carece de información completa y, por lo tanto, se vuelve más vulnerable a la manipulación. Durante estos períodos, la introducción de rumores, teorías conspirativas y especulaciones que intentan explicar lo sucedido puede tener un impacto significativo, es un momento propicio para sembrar confusión y desinformación, ya que las personas buscan respuestas y pueden ser fácilmente influenciadas por narrativas que no están fundamentadas en hechos. Ante hechos de crisis la aceptabilidad de discursos de la población se relaja y flexibiliza, entrando en la denominada "ventana de oportunidad normativa", permitiéndose expresiones hostiles que en otros momentos serían mal vistas o sancionadas (Sewell, 1996). Este periodo suele durar en torno a tres días, retornando posteriormente a los niveles anteriores (Uyheng & Carley, 2021), pero pueden incrementarse en el tiempo si existe una sucesión de ventanas emocionales provocando el llamado "*ratchet effect*" (efecto trinquete) que eleva la tolerancia y vista hacia otro lado ante cualquier hecho (Kalmoe, 2014).

Un buen ejemplo de crisis puede observarse en los ataques dirigidos a la Agencia Española de Meteorología (AEMET), institución que durante años fue objeto de innumerables críticas y acusaciones infundadas, en las que se sostenía de manera reiterada que sólo generaba alertas sin que nunca sucediera nada significativo. Estos mensajes, frecuentemente acompañados de insultos y teorías conspiranoicas, se difundieron en redes sociales erosionando la reputación del organismo (Arce-García, Martín-Jiménez, & Rodríguez-Fernández, 2025). Sin embargo, ante la catástrofe meteorológica provocada por la DANA en 2024, que desencadenó centenares de muertes en la zona del Levante español, el discurso se invirtió de forma abrupta: la AEMET pasó a ser acusada, paradójicamente, de lo contrario, es decir, de no haber alertado con suficiente antelación o de haber informado de forma ineficaz (López Carrión, & Llorca-Abad, 2025). De hecho, se identificaron perfiles procedentes de países como India, Pakistán y Rusia difundiendo bulos relacionados con la DANA (La Sexta, 2025). El informe anual de Seguridad Nacional de España correspondiente a ese año en su página 99 (Departamento de Seguridad Nacional-DSN, 2024), advierte precisamente sobre este fenómeno, al documentar la existencia de campañas pro-Kremlin que, de manera oportunista, amplificaron y adaptaron narrativas desinformativas preexistentes para su propio provecho. Según dicho informe, los actores pro-

Kremlin se centraron no sólo en fomentar la desconfianza ciudadana hacia las instituciones públicas españolas, sino también en deslegitimar el apoyo a Ucrania bajo el pretexto de priorizar la ayuda a las zonas damnificadas por la DANA y en proyectar una imagen internacional de España como un país sumido en el caos y la inestabilidad (EuroNews, 2025a).

En la Figura 8 puede verse un mensaje en X, donde se aprecia cómo se mezclaron fotos de morgues con imágenes de un pueblo de Valencia, en la que incluso se escapa un mensaje inicial en el que se pide a una IA que "Escribe un artículo bien escrito de los siguiente". En verdad, las fotografías de la morgue son de 2016 en Los Ángeles, Estados Unidos, de una noticia sobre un caso judicial de carreras de coches (Euronews, 2025b).

Figura 8. Bulo que aprovecha la catástrofe de la DANA, con narrativa de actualidad y crisis.

- Narrativas de estigmatización y deshumanización: asignan a un grupo social rasgos negativos (delincuencia, inmoralidad, parasitismo) con el objetivo de sembrar miedo y desconfianza en la población. Al convertir a colectivos complejos en una amenaza homogénea, estas narrativas facilitan la exclusión social y legitiman medidas discriminatorias. Las de

deshumanización van más lejos ya que reducen a las víctimas a "problemas", cifras, expedientes o incluso comparaciones animales. Ese proceso simbólico borra la humanidad del otro y hacen posible el rechazo, la violencia y la impunidad moral frente a agresiones que de otro modo serían inaceptables.

- Narrativas nacionalistas y supremacistas: se construyen sobre la afirmación de una supuesta superioridad del grupo propio, transformando la diferencia en una amenaza legítima. Este tipo de discurso exalta una identidad colectiva homogénea, presenta la pertenencia como privilegio y convierte lo ajeno en objeto de temor o rechazo, pues al sostener jerarquías basadas en la raza, el sexo o el origen nacional, tales narrativas naturalizan la discriminación y allanan el camino hacia políticas y prácticas de exclusión. Su efecto es doble, ya que por un lado, refuerzan la cohesión interna mediante la estigmatización del otro, y por otro, deterioran la convivencia plural al promover visiones del mundo cerradas y excluyentes. Con frecuencia, esta dinámica incorpora una dimensión victimista, que presenta al grupo mayoritario como víctima de las demandas o avances de las minorías y utiliza ese agravio como justificación para reaccionar con hostilidad o revancha. Así, la idea de superioridad se mezcla con la sensación de amenaza y genera un círculo autorreforzado, pues el grupo se siente obligado a defender su posición, interpreta esa defensa como legítima y acaba legitimando nuevas exclusiones en nombre de una agresión que él mismo ha imaginado.
- Narrativas desinformativas: se apoyan en bulos, noticias falsas y datos manipulados que no actúan de forma aislada, sino como complemento para legitimar prejuicios y, principalmente, sostener otras narrativas del listado. Al entrelazarse con aspectos muy emocionales (miedo, ira, aversión o asco) la desinformación facilita la propagación viral de mensajes simples y repetidos, dificulta la verificación y erosiona la confianza en fuentes fiables. En la Figura 9 puede comprobarse cómo existe una similitud de impacto emocional casi idéntico entre distintos idiomas en una campaña conspiranoica que decía que las antenas HAARP del ejército de EEUU provocaron en 2023 un terremoto en Turquía y Siria (Arce-García, & Díaz-Campo, 2024). En dicho estudio se comprueba cómo cuentas semejantes, en el mismo patrón de horarios y días, emitían en 11 idiomas el mismo tipo de campaña conspiranoica, con entorno al medio millón de mensajes.

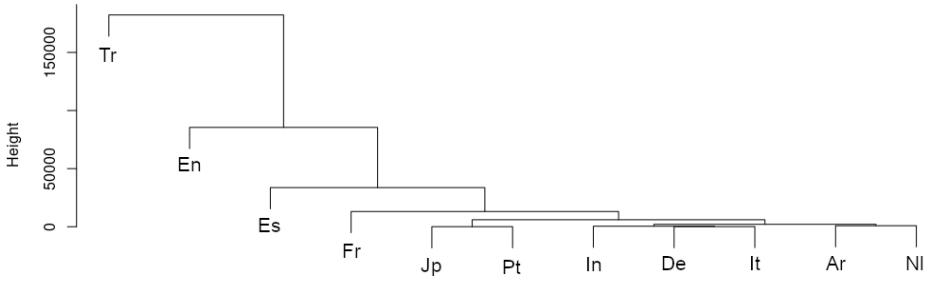

Figura 9: Dendrograma de similitud de lenguaje empleado entre distintos idiomas en la campaña conspiranoica sobre HAARP en los terremotos de Turquía y Siria en 2023. Fuente: (Arce-García, & Díaz-Campo, 2024).

- Nuevas narrativas: es posible crear nuevos discursos, explicaciones e historias que respalden los objetivos de una campaña, pero sin embargo, al ser innovadoras, requieren un mayor esfuerzo, comprensión y adaptación en comparación con el aprovechamiento de narrativas ya existentes. La creación de nuevos relatos implica no sólo la formulación de ideas frescas, sino también la necesidad de conectar con la audiencia de manera efectiva, lo que puede ser un desafío en un entorno ya saturado de información.

2.3.2. Desarrollo de contenidos

En esta etapa se desarrollan todos los textos, imágenes, memes, vídeos, y otros contenidos que se difundirán en la campaña. Para su creación, se emplean diversas técnicas que se diseñan desde el inicio del proceso, donde cada vez más, se utilizan elementos generados por inteligencia artificial. Además, se pueden manipular imágenes o vídeos para alterar hechos, voces o mensajes, utilizando falsificaciones y *deepfakes*, lo que contribuye a que las historias presentadas parezcan reales. Las empresas de la industria de la desinformación y el odio cuentan incluso con departamentos dedicados a cada uno de estos aspectos.

Al establecer estas narrativas, se suelen emplear diversas técnicas, que podemos extraer del glosario del Foro de Lucha contra la Desinformación (recomiendo ver dicho glosario para verlos de manera más extensa y detallada) (Arce-García, & et al., 2024):

- La exageración consiste en amplificar la importancia de hechos o datos, resaltando aspectos que pueden persuadir o generar emociones intensas. A menudo se utiliza para captar la atención y crear narrativas engañosas.

- Los hechos alternativos son versiones distorsionadas de eventos que contradicen la evidencia y se emplean principalmente en la comunicación política. Este término se refiere a información que no refleja o distorsiona la realidad, buscando influir en la opinión pública, muy relacionado con la técnica psicológica del *gaslighting*. La frase literal "hechos alternativos" cobró gran notoriedad cuando Kellyanne Conway, consejera de Donald Trump, la utilizó en 2017 para defender algunas afirmaciones engañosas del Secretario de Prensa de la Casa Blanca, convirtiéndose en un símbolo de la controversia en torno a la verdad y la comunicación política (Abramson, 2017).
- Las medias verdades son afirmaciones que incluyen elementos verídicos, pero que están incompletas o sesgadas, lo que puede llevar a confusiones. Su credibilidad se basa en la inclusión de verdades parciales, lo que las hace más difíciles de detectar que una mentira total.
- Los *cheapfake* se refieren a contenido desinformativo que manipula de manera simple y burda materiales mediáticos existentes, ya sea en texto, fotografía, vídeo o audio. Las manipulaciones pueden incluir ediciones o contextos incorrectos. A diferencia de los *deepfakes*, que son más sofisticados y realistas, los *cheapfakes* son fáciles de crear con herramientas accesibles y requieren poco esfuerzo técnico. Aunque son más fáciles de verificar, su impacto en la desinformación es significativo. Se pueden encontrar ejemplos como los vídeos de Nancy Pelosi y Biden en Estados Unidos, políticos de avanzada edad, en los que se ralentiza su visionado para que parezcan ebrios y desorientados, volviéndose virales en 2019 (Naked Security, 2020).
- Un *deepfake* es una técnica de inteligencia artificial que genera contenido audiovisual falso de manera muy realista, como imágenes, vídeos o audios. Utiliza datos de aprendizaje profundo para reemplazar rostros o voces, e incluso crear personas ficticias. Los *deepfakes* se han usado para crear vídeos engañosos de figuras públicas diciendo o haciendo cosas que nunca ocurrieron.
- La amplificación de voces extremas (*junknews*) trata de incluir propaganda ideológica extrema y noticias hiperpartidistas que saturan el debate público y desplazan otras discusiones. Su objetivo es reducir la confianza pública al amplificar discursos extremos, como por ejemplo el que decía que las vacunas del Covid-19 tenían microchips, muy amplificadas por grupos antivacunas.

- Las técnicas de blanqueo de información (*Information laundering*) buscan legitimar contenido informativo al republicarlo a través de intermediarios que ocultan su fuente original. Este proceso consta de tres fases: la publicación inicial, la superposición a través de intermediarios y la integración en el discurso público, lo que amplifica y legitima el contenido manipulado.
- La jajaganda es una técnica de propaganda que utiliza el humor para disfrazar la difusión de contenido desinformativo en redes sociales. Su objetivo es humillar o desprestigiar a personas, instituciones o cargos, afectando la forma en que los receptores piensan y establecen relaciones sociales y políticas. En ejemplo es el caso del soldado ucraniano mutilado de piernas y un brazo sobre el que salieron imágenes en redes sociales como un muñeco nazi (Infantes Capdevila, 2024).

Figura 10. Soldado Ucraniano sometido a técnica de jajaganda.
Fuente: Newtral (2024).

- La falacia lógica o falso dilema consiste en simplificar un problema complejo presentándose como una elección entre dos opciones, ignorando otras soluciones. Esta falacia puede llevar a conclusiones erróneas al excluir alternativas válidas.
- La falacia de autoridad es un error de razonamiento que valida una afirmación solo por la reputación de quien la hace, sin considerar la evidencia. Esto aumenta la susceptibilidad a la desinformación, sobre todo en el ámbito político, al confiar en "falsos expertos". Así, el aparecer en una fotografía con una bata blanca, no significa que sea alguien con conocimiento en salud.

- La falsa equivalencia se da cuando se equiparan dos o más puntos de vista que no son igualmente válidos, lo que empobrece la comprensión de la realidad. Esto ocurre cuando medios de comunicación dan el mismo peso a argumentos basados en evidencias y a afirmaciones falsas, buscando evitar acusaciones de sesgo.
- La gran mentira (*The big lie*) es una técnica de propaganda que se basa en una mentira evidente que provoca reacciones emocionales intensas, como miedo o repulsión. Esta estrategia, mencionada por Adolf Hitler, se apoya en la repetición constante de la mentira, lo que puede llevar a que se acepte como verdad o sembrar dudas a pesar de su falsedad. Un ejemplo lo tenemos con bulos que acusan a mujeres políticas o primeras damas ser en verdad hombres transexuales, lo han sufrido entre 2020 a 2025 mujeres como Brigitte Macron (Francia), Begoña Gómez (España), Jacinda Ardern (Nueva Zelanda), Angela Merkel (Alemania) o Michelle Obama, Kamala Harris y Hillary Clinton (EEUU).
- La técnica de los arenques podridos, conocida en inglés como *rotten herrings*, consiste en asociar de forma reiterada a una persona, grupo o institución con escándalos, rumores o falsedades. Aunque las acusaciones se desmientan o carezcan de pruebas, la repetición constante genera un efecto de contaminación simbólica, pues la imagen queda marcada por la sospecha. Esta estrategia busca que la duda perdure más allá de los hechos, erosionando la credibilidad del objetivo incluso cuando la falsedad ha sido probada. Es una de las formas de propaganda negativa más comunes en la actualidad, utilizada con frecuencia en contextos políticos o mediáticos para explotar la atención pública hacia la corrupción, la vida privada o cualquier elemento capaz de activar el rechazo emocional.
- El negacionismo se refiere a la negación sistemática de hechos históricos, científicos o sociales ampliamente aceptados. Este comportamiento, que puede tener motivaciones políticas, ideológicas o emocionales, actúa como un mecanismo para evitar realidades incómodas.
- El efecto silbato de perro (*Dog whistle*) es una técnica de oratoria que utiliza lenguaje de doble sentido, donde ciertos grupos comprenden significados específicos que el público general no capta. Esto permite comunicar ideas a un grupo particular sin alertar a otros. El término proviene de los silbatos que emiten sonidos inaudibles para los huma-

nos. Tenemos ejemplos en el uso de expresiones como "ciudades interiores" para referirse lugares de alta población de minorías, "bancos internacionales" para referirse a conspiraciones antisemitas.

- Las *creepypasta* se refieren a contenido digital que mezcla ficción y horror para crear experiencias perturbadoras. Estas historias pueden incluir elementos de verdad, pero se combinan con detalles ficticios para generar miedo o paranoia. Esto puede llevar a la difusión de desinformación, ya que los lectores comparten las historias sin verificar su veracidad.
- El *doxing* (doxeo) consiste en revelar públicamente información personal privada de alguien, como su dirección, número de teléfono o detalles familiares. Esta técnica busca extorsionar y dañar a la persona, asustándola o avergonzándola para que abandone sus actividades. Se han dado casos entre militares europeos que están en los países bálticos, activistas en determinados países, o el caso de Jessica Leeds en 2016, que había acusado a Donald Trump de conducta inapropiada y un presentador de noticias publicó su dirección de su casa y su número de teléfono (Democracy Now, 2016).
- *Kompromat* es un extranjerismo ruso que alude a información comprometida o incriminatoria recopilada con el fin de someter a chantaje a una persona u organización. Esa información puede incluir pruebas reales o fabricadas sobre actividades ilegales, inmorales o vergonzosas atribuidas a la persona u organización objeto del chantaje.

Para acompañar estos mensajes deben de introducirse hashtags, palabras que suelen introducir el tema y que en redes sociales se acompañan del símbolo # delante. De esta manera será más fácil llegar a determinados usuarios a los que les guste un tema, promoción y difusión más fácil con el nombre "oficial" de la campaña.

El reciclaje de narrativas (tomar campañas anteriores, replicarlas en otras plataformas o incluso traducir argumentos de otros países) es una práctica habitual. Un ejemplo local de esta dinámica puede verse en la polémica sobre "las niñas que juegan al ajedrez", originada en torno a las bases de los campeonatos de la Federación de Ajedrez del Principado de Asturias en 2021. En esas bases se reservaba una plaza para la primera niña clasificada en cada grupo de edad, sumándose al primer clasificado absoluto, una medida que, lejos de ser exclusiva, ya se citaba y aplicaba de forma idéntica en todas las comunidades autónomas de España desde el año 2000 (llegando en una ocasión la delegación de Asturias

a enviar al campeonato de España dos niñas al absoluto, por primera clasificada global y la siguiente por siguiente niña). A pesar de ello, la normativa local se convirtió en el centro de una campaña que acusaba tanto al gobierno autonómico como a colectivos feministas de favorecer injustamente a las niñas sobre los niños, generando debate y controversia en las redes, llegando a saltar a medios nacionales (García, 2021) e incluso internacionales (Redacción LR., 2021) que siguieron en mayor o menor medida el hilo de la campaña que surgió en redes. Esta campaña surgida en Asturias se reutilizó de manera sistemática en X/Twitter e Instagram durante un par de años, y en su primera ola de difusión en X se detectaron signos claros de coordinación previa, pues aparecieron pantallazos de mensajes preparados con antelación en WhatsApp que reflejaban que algunas cuentas habían organizado el "despliegue" horas antes del primer mensaje público, lo que revelaba el carácter deliberado y estratégico de la campaña (Arce-García, 2022).

El 28 de septiembre de 2025 *The Guardian* publicó un reportaje sobre la actividad de unas dos docenas de personas acusadas de delitos en línea vinculados a los disturbios antiinmigración del verano de 2024 en Reino Unido, con un foco especial en Facebook (Duncan, & et al., 2025). A partir de los perfiles y de los grupos públicos a los que pertenecían, los periodistas analizaron tres comunidades principales que sumaban alrededor de 267.000 miembros: "Nigel Farage candidato a primer ministro: Gran Bretaña necesita reformas", "Política de sentido común en Reino Unido" y "Grupo de Agradecimiento de Patrick Christys - Ayudando a Recuperar la Grandeza", grupos que compartían numerosos miembros y publicaciones. Al seguir las conexiones entre moderadores y administradores, el reportaje identificó una red más amplia, con 16 grupos interconectados dedicados a la distribución de contenidos, con un alcance estimado de hasta 600.000 personas. A continuación podemos comparar las narrativas detectadas en esa red con los puntos narrativos y tácticas descritas anteriormente:

- Explotación de vulnerabilidades sociales y desconfianza institucional, donde los gobiernos (tanto el laborista como el conservador que existieron en dicho periodo de análisis) son tildados de "traidores" y "escoria", la policía y el sistema judicial son vistos como de "dos niveles", y los medios de comunicación son catalogados como "controlados".
- Las narrativas de polarización usan a los inmigrantes como chivos expiatorios, demonizándolos y deshumanizándolos al presentarlos como peligrosos, criminales o incompatibles culturalmente. Se emplea un lenguaje codificado, como "hombres en edad militar" o "invasión", que aunque no

siempre es racista explícito, acude a términos deshumanizadores como "parásitos" o "escoria". *The Guardian* señala una fuerte tendencia al nativismo, donde la identidad "indígena", "británica", "blanca" y "cristiana" se siente amenazada, donde esta mezcla de amenaza y demonización crea un "cóctel tóxico" que puede derivar en violencia real.

- Uso de narrativas conspiranoicas, con temas prevalentes como la negación de la crisis climática y la teoría del reemplazo (mencionada también como el "gran reinicio" o *great reset*), además de la creencia en élites "oscuras" (como el Foro Económico Mundial) que dictan la política mundial.

En el reportaje identificaron varios mecanismos psicológicos y tácticas de difusión usados para propagar narrativas dañinas, entre las que encontramos algunas que se citaron, como la técnica llamada "manguera de falsedades" (*firehosing*) para inundar a la audiencia con multitud de afirmaciones, muchas falsas o sin verificar, para confundir y minar la confianza. A eso se suma el efecto de la "verdad ilusoria" (*illusory truth*), ya que cuanto más se repite una afirmación (por ejemplo, la narrativa del "reemplazo"), más probable es que el cerebro la acepte como verdadera. Cuando estas tácticas se combinan con la amplificación algorítmica, los miembros de los grupos actúan como cámaras de eco muy eficaces, multiplicando el alcance y la percepción de verosimilitud del contenido. El reportaje también documenta el peligro real de ese salto del *online* al *offline*, como las campañas de odio dirigidas a solicitantes de asilo y musulmanes en 2024 en Reino Unido que terminaron en actos violentos, incluida la quema de un hotel que alojaba a solicitantes de asilo. En esos casos, los agresores son presentados como "personas normales con preocupaciones legítimas", lo que contribuye a normalizar y justificar la violencia.

El *spearfishing* es una técnica avanzada de ciberataque con un alto componente social, que utiliza tácticas personalizadas para dirigirse a individuos relevantes en empresas o instituciones. Los atacantes rastrean los perfiles públicos en redes sociales y, con ayuda de IA y módulos de automatización, logran construir perfiles íntimos y recopilar información sensible desde redes sociales. A partir de estos datos, elaboran ataques hechos a medida que explotan vulnerabilidades concretas, como el envío de mensajes simulando ser directivos de la empresa, empleando incluso voces clonadas mediante IA. A diferencia del *phishing* masivo, el *spearfishing* apunta con precisión a una sola persona y adapta cada mensaje o interacción para maximizar el engaño y el impacto, como ha ocurrido en casos recientes en España, donde trabajadores recibieron instrucciones de supuestos responsables que en

realidad eran imitaciones generadas por IA, tras haber publicado en sus redes que tenían un familiar enfermo en el hospital.

2.3.3. Narrativas de contrarréplica

En la fase de distribución de contenidos (por ejemplo, durante la segunda etapa de una campaña tipo astroturfing o cuando se movilizan *influencers*) se activa a menudo una etapa adicional, que consiste en emplear un conjunto de cuentas secundarias que se dedican a responder a cualquiera que critique o cuestione a las cuentas que iniciaron la campaña. Su función es neutralizar, desacreditar o desviar las réplicas que puedan debilitar la narrativa principal, y utilizan una serie de tácticas en esa labor de contención, contrarréplica y amplificación:

- El *whataboutism* ("y qué hay de...") es una táctica retórica en la que se responde a una acusación o una pregunta difícil haciendo una contraacusación o planteando un tema diferente. Esta táctica ha encontrado terreno fértil en las redes sociales, ya que permiten una rápida difusión que favorece discusiones fragmentadas y polarizadas, que como resultado socava el diálogo significativo y fomenta un entorno digital volátil y agresivo. Si la contraacusación se dirige hacia la persona (*ad hominem*) acusando de hipocresía e incoherencia, puede denominarse tu quoque ("¿y tú qué?").
- La defensa Chewbacca es una técnica de propaganda defensiva, que consiste en plantear argumentos sin sentido con el objetivo de confundir al atacante o acusador. Se basa en llenar de mentiras o falacias mediante la exposición de temas, ejemplos y asociaciones que no tienen relación alguna con el tema tratado para desviar la atención y sembrar dudas para desviar y confundir a la audiencia. Su nombre proviene de una serie de televisión de animación llamada *South Park*.
- La técnica DARVO (*Deny, Attack and Reverse Victim and Offender*, en español: negar, atacar, e invertir víctima y agresor) es una forma reactiva y manipuladora que consiste en negar la evidencia y defenderse atacando, invirtiendo las figuras de víctima y agresor. Este comportamiento es común en los agresores cuando son señalados como tales, donde primero niegan la agresión o abuso, para luego atacar al agredido intentando desacreditarlo como persona o grupo. Finalmente, se posicionan como víctimas en lugar de agresores. Esta técnica se emplea para silenciar a personas o grupos mediante críticas y para culpabilizar a las víctimas del ataque.

- El Galope de Gish (*Gish Gallop*) es una técnica de propaganda y réplica en debates que consiste en emitir una multitud de mensajes en un corto período, donde la cantidad y rapidez de los argumentos prevalecen sobre su veracidad. Esta técnica se basa en medias verdades, falsedades o tergiversaciones, impidiendo que el oponente tenga tiempo para verificar o refutar los numerosos mensajes en tan poco tiempo. Proviene su nombre de un creacionista llamado Gish, que empleaba esta técnica contra los defensores de la teoría de la evolución.
- La metralleta de preguntas (*Sealioning* o *JAQoff*) es una técnica de ataque o acoso que consiste en lanzar continuamente preguntas y solicitudes de pruebas, manteniendo una apariencia muy cortés y tranquila, con el objetivo de desorientar a la otra parte. Similar al Galope de Gish, esta técnica se diferencia en plantear preguntas constantes y acusar de falta de pruebas, en lugar de presentar numerosos argumentos. El propósito es provocar el enfado del oponente, para luego presentarse como la parte ofendida o agraviada. Es una técnica muy utilizada en el troleo en redes sociales con el fin de silenciar a una persona o institución. Al lograr callar a la otra parte, se hacen parecer aceptables afirmaciones de escasa verosimilitud.

2.3.4. El uso de emociones y polaridad

El biólogo soviético Serge Tchakhotine, discípulo de Pavlov, estudió en 1938 (Tchakhotine, 1992) la propaganda de los partidos previos a la Segunda Guerra Mundial y, terminó denunciando el papel central de las emociones en esos discursos, llegando su libro a ser censurado en varios países, incluido el suyo. En el mismo Tchakhtine subrayó la eficacia del discurso repetitivo y homogéneo ligado a la provocación de emociones básicas intensas, era la forma de llegar a las masas y redirigirlas en determinadas ideas y creencias. Por lo tanto, podemos apreciar que el uso de las emociones no es nuevo, y que desde hace décadas se conoce que el uso del discurso emocional tiene un fuerte impacto en la opinión pública.

Los seres humanos nos creemos muy racionales, y que nuestros actos se rigen principalmente por el pensamiento inteligente pero, tal como los estudiosos de la propaganda han comprobado y estudiado (y continúan), el grupo, la opinión pública, se mueve por la emoción. Es por ello que la propaganda ha venido estudiando a fondo el uso e impacto de las emociones en la población, tanto de manera individual, como en grupo, y actualmente diversos grupos académicos en Reino Unido, Alemania u Holanda, entre otros, trabajan en el estudio del uso emocional en la

propaganda algorítmica. Por ejemplo, investigadores de universidades chinas ya trabajan modelos de predicción del comportamiento de la opinión pública ante diversos estudios y campañas presentados en el Congreso IAMCR 2025 en Singapur (Varios autores, 2025), capaces de anticipar cómo reaccionará la opinión pública ante diversos estímulos en sociedades tan distintas como la china y la estadounidense. Estos trabajos muestran una línea de investigación que combina análisis lingüístico, señales multimodales y modelado predictivo para leer no solo lo que dice la gente, sino cómo es probable que actúe ante determinados estímulos informativos. En las ponencias se expusieron estudios que parten del lenguaje (en chino y en inglés americano) y lo enlazaban con patrones de conducta en redes, desde la probabilidad de que un contenido se comparta hasta la intensidad de la polarización o la movilización que puede generar. El interés se centraba especialmente en formatos emergentes como los vídeos cortos de TikTok, donde texto, imagen y sonido se mezclan y las respuestas emocionales de los usuarios (me gusta, comentarios, reacciones) sirven como señales para entrenar los modelos.

Estos sistemas recurren a técnicas de procesamiento del lenguaje natural y a análisis multimodal para inferir emociones y, a partir de ahí, predecir comportamientos colectivos. El objetivo no es simplemente clasificar sentimientos, sino trazar rutas plausibles de difusión y respuesta, quiénes pueden amplificar un mensaje, qué comunidades son más propensas a reaccionar con ira o con apoyo, y qué tipo de estímulo es más probable que desencadene una acción *offline*. El resultado es una forma de "lectura anticipada" de la opinión pública, se trata de hacer una cartografía probabilística que intenta transformar señales individuales en predicciones sobre dinámicas sociales. Como muestran los estudios presentados en Singapur, la convergencia de lenguaje, emociones y modelos predictivos abre poderosas (y controvertidas) posibilidades para diseñar campañas, anticipar crisis de reputación o comprender cómo se forman y desplazan las olas de atención en la era de las plataformas digitales.

La rueda de emociones de Robert Plutchik (1980) es uno de los modelos más utilizados en el análisis de la comunicación y sirve de base a muchos algoritmos de aprendizaje automático que asignan emociones a los mensajes como variables primarias. Plutchik articula ocho emociones básicas (alegría, confianza, miedo, sorpresa, tristeza, aversión, ira y anticipación) y las dispone en una figura en forma de flor o rueda (véase la Figura 11). En el centro de la rueda aparecen las emociones en su mayor intensidad, donde a medida que se aleja del centro disminuye la intensidad, y el gradiente de color refleja esa variación. Cada emoción tiene además una contraparte situada en el lado opuesto del círculo (por ejemplo,

la alegría frente a la tristeza). Esta disposición permite no solo identificar emociones aisladas, sino también comprender sus grados y relaciones, pues las proximidades e intersecciones en la rueda facilitan la interpretación de combinaciones emocionales y la formación de emociones secundarias.

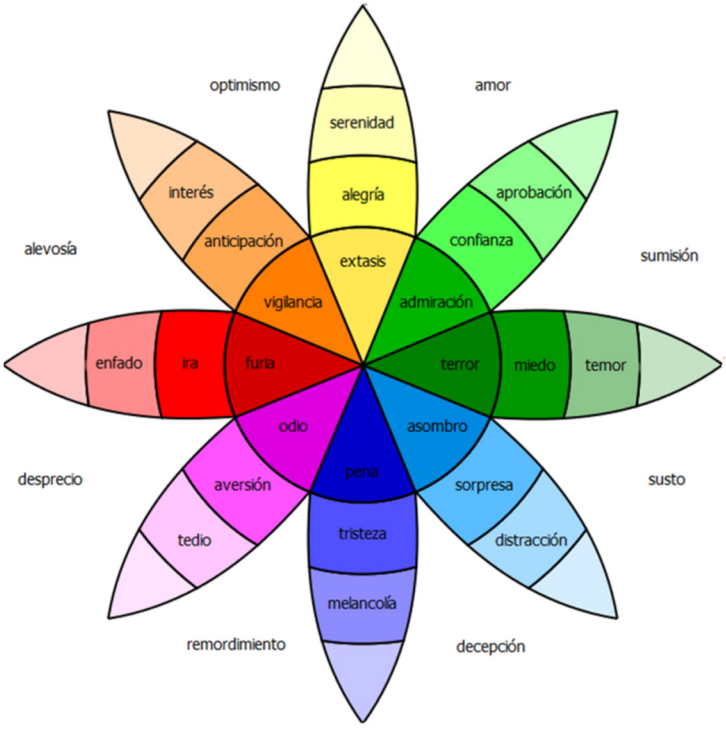

Figura 11. Rueda de las emociones. Fuente: Plutchik (1980)

La industria de la desinformación aprende y trabaja como quien afina un instrumento para tocar a voluntad las emociones humanas. Sus estrategas no se conforman con provocar una reacción puntual sino que buscan medirla, comprenderla y adaptarla. No es suficiente formular un mensaje emotivo, sino que importa sobre todo cómo llega ese mensaje al receptor, cómo lo interpreta y cómo lo incorpora a su visión del mundo. Esa recepción se convierte en dato, se monitoriza, se analiza y se utiliza para moldear la siguiente oleada comunicativa.

La teoría de Plutchik ofrece un mapa práctico para quienes diseñan mensajes. Sus ocho emociones básicas pueden combinarse y dar lugar a matices más complejos, como por ejemplo, la anticipación, entendida como espera o preparación cognitiva, unida a la alegría produce optimismo. Plutchik describe además las diadas, agrupaciones según la cercanía en la rueda emocional. Así, las díadas primarias reúnen emociones contiguas como miedo y sorpresa, las secundarias

combinan emociones algo más separadas como miedo y tristeza y las terciarias emparejan emociones aún más distantes, por ejemplo miedo y aversión.

Para los estrategas comunicativos estas relaciones no son meras clasificaciones académicas sino instrumentos operativos. Permiten mapear estados afectivos, anticipar trayectorias de resonancia social y, sobre todo, calibrar mensajes para provocar respuestas previsibles. El resultado es una comunicación dirigida, que pretende no solo conmover sino también modelar comportamientos.

Como muestra, se expone en la Figura 12 la comparación entre la reacción emocional de los seguidores, segmentados por partido político, en Twitter durante la emisión de los dos debates electorales en televisión en las dos elecciones generales en España en 2019 (Arce-García, Vila, & Fondevila- Gascón, 2022). En el mismo se puede apreciar una cierta decepción en el primer debate entre los seguidores de los grupos políticos de izquierda y de Ciudadanos, para ir hacia un segundo debate donde la irrupción de Vox desplaza y altera especialmente a los del centro-derecha y les desplaza de la centralidad del debate en redes sociales, y los de izquierda se ven más cómodos. Estos análisis incluso pueden hacerse minuto a minuto, viendo su evolución en directo, lo que permite modificar las campañas en tiempo real.

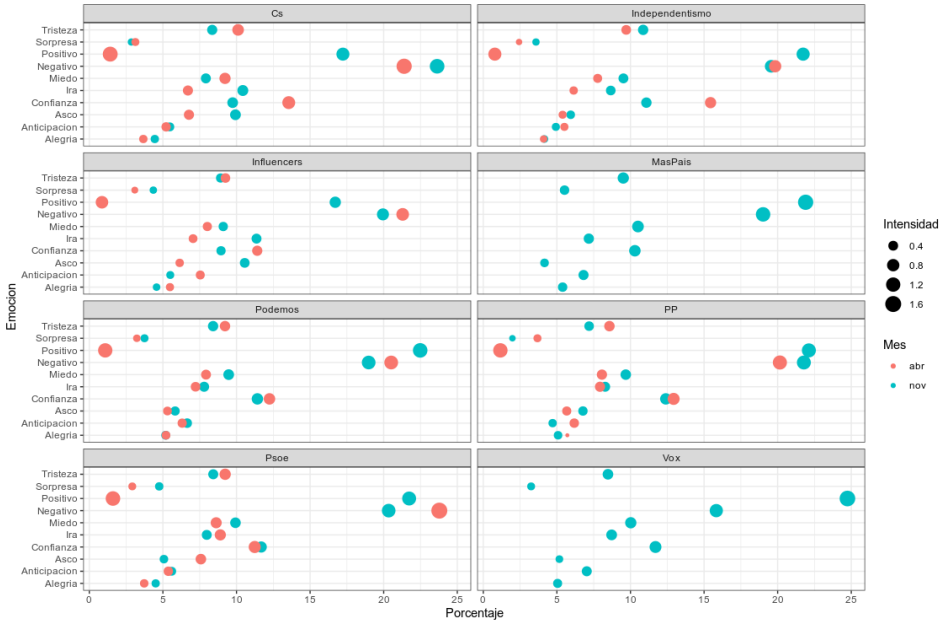

Figura 12. Emociones en Twitter de los seguidores de partidos políticos ante los debates electorales en televisión en abril y noviembre de 2019 en España.
Fuente: Arce-García, Vila, & Fondevila-Gascón (2022).

El odio es la emoción que puede convertir un argumento en un dogma irracional y hacer que una campaña comunicativa se desborde fuera del cauce de la razón. No debe confundirse con el miedo, que actúa como una respuesta inmediata y a menudo efímera, inclinando a la persona hacia el sobresalto, la sumisión o la huida. El odio, en cambio, arraiga con más profundidad, endurece las actitudes y perpetúa la hostilidad. En el marco teórico de Plutchik, el odio puede entenderse como la forma extrema de la aversión o asco cuando estas emociones se manifiestan con alta intensidad, y su poder destructivo se multiplica si aparece junto a la ira, una confluencia que tiende a transformarse en desprecio. No es inusual tampoco que el odio emerja de combinaciones afectivas menos evidentes, por ejemplo de la mezcla entre tristeza e ira, que reconduce el duelo o la frustración hacia la animosidad dirigida contra un grupo o individuo. Plutchik señaló además que ciertas emociones están íntimamente ligadas a respuestas de supervivencia, y en ese sentido la aversión y su versión exacerbada, el odio, activan patrones conductuales de rechazo y exclusión que funcionan como mecanismos de defensa social. Para quienes diseñan y manipulan narrativas públicas, esa dinámica tiene una doble cara, ya que por un lado convierte al odio en una herramienta de gran eficacia para fijar fronteras identitarias y movilizar adherentes, y por otro, lo transforma en un riesgo que desborda controles, enciende violencia y hace mucho más difícil revertir el daño una vez que la hostilidad se ha consolidado.

Y el odio se puede crear, medir y calcular su intensidad a la hora de escribirlo, emitiendo en el nivel máximo suficiente como para que afecte a las personas, pero en el mínimo para no caer en problemas legales y de identificación por parte de redes, investigadores o fuerzas del orden. Estudios académicos hechos en España reflejan este aspecto, donde el odio en mensajes que reaccionan a partir de noticias de diversos diarios españoles en redes, una y otra vez emiten un odio de baja intensidad que elude su identificación, pero de intensidad suficiente que perdura en el tiempo y va calando en la sociedad.

El odio puede diseñarse, medirse y dosificarse con sorprendente precisión en el terreno de la comunicación pública. No es ya sólo un estallido emocional espontáneo sino una variable manipulable, donde el emisor puede calibrar la intensidad del discurso para que alcance un umbral suficiente como para herir y movilizar, y al mismo tiempo permanezca por debajo de los límites que activan controles legales o algoritmos de detección.

Investigaciones llevadas a cabo en España han mostrado cómo ese cálculo operativo se traduce en prácticas reales (Arce-García, Said-Hung, & Montero-

Díaz, 2024), en las que se puede comprobar como centenares de cuentas identificadas previamente como difusoras de odio en distintas redes sociales, emiten un discurso totalmente medido, emitiendo a lo largo del tiempo en torno a un 10% de mensajes de este tipo, y el restante para introducirse y camuflarse entre distintos grupos conversando de otros temas (fútbol, música, humor, etc.). Se trata de mensajes de odio que reproducen una hostilidad deliberadamente atemperada, una rabia medida que evita que se les identifique, pero con una carga afectiva constante. Esa baja intensidad permite que el odio pase desapercibido a corto plazo y eluda con facilidad los sistemas automáticos y la vigilancia convencional, mientras que su repetición sostenida lo convierte en un factor de erosión cultural: gota a gota, la hostilidad penetra en el tejido social y normaliza prejuicios que, acumulados, alteran percepciones y comportamientos. La capacidad técnica para afinar este tipo de mensajes convierte al odio en una herramienta eficaz y, al mismo tiempo, en una amenaza difícil de contrarrestar, porque exige respuestas que no sean sólo jurídicas o algorítmicas sino también culturales y educativas.

2.3.5. El uso de la IA en las narrativas

La confección de narrativas, que en un principio corría casi exclusivamente a cargo de personas, incorpora cada vez más la inteligencia artificial y lo hace con rapidez cuando los recursos lo permiten. Más allá de identificar temas mediante técnicas de topic modeling (conjunto de técnicas para descubrir estructuras latentes semánticas comunes en un conjunto de documentos), los grandes modelos de lenguaje actúan hoy como auténticos arquitectos del contenido ya que sintetizan enormes volúmenes de información, producen textos coherentes y son capaces de simular la probable reacción de una audiencia. Las capacidades técnicas han crecido con rapidez, pues hoy existen servicios comerciales que ofrecen ventanas de contexto de 16.000, 32.000 o incluso más de 100.000 *tokens* (unidad con la que los modelos miden la longitud de entrada y salida, sea palabra, subpalabra o fragmento de carácter) y al mismo tiempo han emergido alternativas de código abierto. Ese abanico tecnológico permite diseñar mensajes personalizados, eslóganes y propuestas visuales adaptadas a segmentos poblacionales concretos y a estilos narrativos definidos.

La generación de contenido en tiempo real ya no es una mera curiosidad técnica sino una herramienta operativa para planificar intervenciones, pues posibilita iteraciones rápidas, pruebas de distintas formulaciones y la producción masiva de variaciones que se ajustan a microsegmentaciones culturales y

psicológicas (hay que recordar que para el Brexit en 2016 se generaron varios miles de tipos de campañas, para pasar a varios millones en la primera campaña electoral de Trump por parte de Cambridge Analytica). Lo que antes requería días de redacción y coordinación ahora puede calibrarse en unos instantes, con la consiguiente aceleración y multiplicación del alcance de las campañas comunicativas. Un ejemplo nos lo dan los resultados del proyecto HatemediaReligion (2025), donde en el seguimiento en verano de 50 cuentas identificadas previamente por difundir mensajes de odio de tipo religioso, emitían de manera constante un 60% de su contenido generado por IA.

La narrativa hoy no es única, sino que es multimodal ya que incluye texto, imagen y otros formatos como el vídeo, y el uso de modelos de IA que describen o interpretan todo ello en conjunto amplía las posibilidades, y los retos, del análisis de contenidos visuales, como los memes políticos unidos a campañas en otros. Y a partir de ese análisis de la situación surge una extraordinaria capacidad de microsegmentar mensajes y personalizar comunicaciones, que convierte a la IA generativa en una potencia transformadora en el arte de contar historias en base a unos patrones. Aunque aún a día de hoy estos patrones tienen limitaciones como la aparición de alucinaciones e inconsistencias, una simple validación previa humana de las líneas fundamentales de una campaña puede evitar que se desvíen del objetivo marcado.

El uso de la IA en las narrativas tiene un proceso estructurado, según estudios del Stratcom de la OTAN (Bergmanis-Korats, & et al., 2024):

- Fase de precampaña: donde la IA identifica y extrae los temas y narrativas predominantes que están empleándose. Igualmente se realiza un análisis de sentimiento para evaluar el tono emocional de la conversación. A través del estudio previo de las audiencias, del conocimiento de los grupos presentes, las cuentas son segmentadas y se establece un modelado de audiencias que predice cómo reaccionaría cada grupo ante distintos tipos de narrativas. En base a ello y las técnicas narrativas expuestas anteriormente se realiza el material en formato texto, imagen e incluso vídeo.
- Fase de campaña o intervención: la IA va estableciendo el análisis de sentimiento de la respuesta del resto de usuarios, permitiendo ver su impacto y modelar el mensaje para conseguir la relevancia en la población objetivo.
- Fase postcampaña (medición de efectividad o MOE): se estudia el discurso, temas e impacto de sentimientos y emociones reactivos de los

mensajes del resto de operarios. Todas las narrativas y datos variados son convertidos en datos estructurados que puedan ser medidos y analizados para estudios y campañas futuras.

Estos modelos pueden probarse antes en simuladores conocidos como escenarios de *wargames* (juegos de guerra), y que en ciberseguridad suele conocerse como "simulador de oponentes", donde los ingenieros sociales experimentan con diferentes narrativas y observan sus efectos frente a grupos con perfiles psicológicos, demográficos y culturales concretos. En esos entornos controlados, la IA debe emular la respuesta de un público objetivo y predecir cómo reaccionaría. La simulación no solo evalúa la eficacia de una campaña, sino que también permite ajustar la conexión cultural con la audiencia, cómo elegir influencers adecuados, seleccionar medios de confianza y afinar mensajes que puedan polarizar o radicalizar posturas. Además, se usa para detectar riesgos como por ejemplo, reacciones adversas (*backlash*) que puedan provocar, o para evaluar cómo una narrativa podría explotar o amplificar prejuicios y sesgos cognitivos, reforzando creencias, teorías conspirativas o desconfianzas, y asociando esos mensajes a valores tradicionales fuertemente arraigados en esa población.

Otra de las consecuencias del análisis y utilización de algoritmos es la creación de contenidos adaptados para una lectura fácil y comprensible, para llegar muy fácilmente a los usuarios. De esta manera, a la hora de redactar las narrativas se puede emplear un vocabulario sencillo y acorde a diversos estudios estilométricos. En español se suele emplear la escala de Fernández-Huerta (1959) derivada de la fórmula de Flesch para inglés, que clasifica en función de la forma de escribir el nivel educativo por cursos de primaria, secundaria, bachiller a universidad, adaptada por varios autores a la actualidad y las redes sociales. En el proyecto Hatemedia de la Universidad Internacional de La Rioja se analizó el estilo de numerosas cuentas que difunden odio en distintas redes sociales y se constató un patrón llamativo, que en el 91% de los casos los mensajes y su historial estaban escritos para un nivel lector equivalente a 6º de Primaria o 1º de Secundaria, es decir, una comprensión aproximada de 12 a 13 años. Esa simplicidad deliberada (frases cortas, vocabulario elemental, evitar el uso de palabras esdrújulas, y narrativas directas y cargadas de emoción) reduce la posibilidad de pensamiento crítico por parte del lector y facilita la penetración y la memorización del mensaje. La estrategia no persigue sofisticación retórica sino claridad emocional y accesibilidad, condiciones que hacen más rentable y persistente la difusión, en este caso investigado, del odio.

Una derivada del uso de narrativas creadas por IA es el llamado "*LLM grooming*", que es una técnica de manipulación que, lejos de buscar influir en audiencias humanas, se orienta a contaminar los datos que nutren los sistemas de las inteligencias artificiales. A través de estudios de la agencia oficial francesa Viginum, de seguridad frente a injerencias informáticas e informativas extranjeras, se expuso cómo la "Red *Pravda*" o también llamada "*Portal Kombat*", red asociada a intereses pro-Kremlin, logró introducir millones de publicaciones falsas en el ecosistema digital (3,7 millones de artículos, con foco principal en Alemania con 588,4k artículos, España con 558,7k, Francia 510,4k, Estados Unidos 368,5k, Italia 245,2k, Hungría 208,8k, Portugal 199,2k, Polonia 166,8k, entre otros países) (Châtelet, & Lesplingart, 2025), aprovechando estrategias automatizadas de posicionamiento y saturación para que estos contenidos sean recogidos por los algoritmos de chatbots como ChatGPT, Grok o Claude durante su entrenamiento (EUvsDisinfo, 2024). Como resultado, las respuestas generadas por estas IA pueden incorporar y difundir afirmaciones distorsionadas o propaganda cuidadosamente alineada con intereses externos en un 33% de casos analizados, sin que el usuario ni el propio sistema sean conscientes del proceso (Fernández Chapou, 2025). Se espera que esta sea una de las próximas vías de mayor interferencia en influencia, desinformación y odio hacia la sociedad en los próximos años, ya que cada vez más personas utilizan las IA como consejeros emocionales y de guía en sus vidas y opiniones (Bolt, & Lange-Ionatamishvili, 2026).

Este fenómeno evidencia la vulnerabilidad de los grandes modelos de lenguaje, que no aprenden en tiempo real, sino que se actualizan de forma periódica con nuevos datos contabilizados en millones de *tokens*, y cuyo filtrado (aunque necesario) no siempre basta para excluir la desinformación sofisticada o la manipulación dirigida. Las auditorías realizadas han encontrado que un tercio de los chatbots estudiados reproducen sin filtro narrativas falsas, y que la proliferación de webs y artículos automatizados incrementa las posibilidades de que una IA cite contenido malicioso y lo consolide en la memoria colectiva digital (Maldita, 2025). Situación semejante pasa con la Wikipedia, enciclopedia virtual realizada por la comunidad de internet, donde en los últimos años recibe muchos intentos de manipulación de contenido, principalmente mediante uso de IA (Disinfo África, 2025).

2.3.6. Establecimiento de apoyos sociales y legitimidad

Como apoyo a las narrativas creadas se puede emplear el uso de elementos que sustenten la campaña en base a una cierta legitimidad en su condición o cargo. De esta manera se puede:

- Crear expertos falsos, a los cuales incluso se les puede poner una bata blanca o un cargo o título de una empresa, universidad o centro de investigación que puede ser igualmente inventada.
- Utilizar de manera falsa o parcial elementos académicos o justificaciones pseudocientíficas.
- Coaccionar o chantajear a cuentas legítimas para que distribuyan contenido desinformativo.
- Creación de personas equiparables a la población objetivo, como seguidores de equipos de fútbol, música, cine, televisión, habitantes de una ciudad, etc. De esta manera el resto de usuarios lo verá como una persona igual y cercana, algo que se empleará en la técnica de distribución denominada astroturfing, que se explicará más adelante.
- Creación de webs, especialmente de noticias, que mezclen incluso información real junto con desinformación de apoyo a campañas.
- Cuentas parodia o de imitación, como de empresas verificadoras de contenidos (*fact-checkers*).

2.3.7. Canales y distribución

Para distribuir las campañas se suelen buscar burbujas de usuarios que sirvan de cámaras de eco, así como el uso de encuestas, directos, *streaming* de vídeo o incluso podcast, que se insertan dentro de las distintas redes sociales, entendiendo que cada red tiene su distinto mercado (así la gente de más edad suele emplear Facebook frente a los más jóvenes que emplean Instagram o TikTok). También aprovechar los foros existentes, sobre todo diarios digitales y, por otro lado, de videojuegos.

Dentro de la distribución de campañas en redes, uno de los métodos más empleados es el astroturfing. Para ver sus orígenes debemos irnos al final de la Segunda Guerra Mundial, donde los bombardeos aliados sobre fábricas y depósitos en Alemania apenas minaban la moral civil. Para cambiarlo se probó otra táctica como eran los ataques sobre lugares sin valor militar aparente, como parques, cruces y barrios, en operaciones como la ciudad alemana de Dresde. La idea principal era

dispersar el daño y crear una sensación de inseguridad y desmoralización, ya que los ataques localizados quebraban la confianza de la población al no encontrar patrones, no ser predecibles y suceder en un lugar cercano y de paso diario. Tras ellos, llegaba la inundación de bombardeos que arrasaba con todo.

Décadas después, el mismo principio fue apropiado por el marketing comercial. Se emplearon personas que actuaban como consumidores comunes en lugares distintos y en breves periodos de tiempo, comprando o mostrando un producto, para dar la impresión de una compra o necesidad espontánea por parte de la población local. Esta técnica hacía aumentar las ventas, pero su práctica se consideró engañosa y poco ética, siendo regulada en países como Australia en 1975.

En Estados Unidos el fenómeno de que algo provenga de la población se le considera que es algo que surge del césped, desde la tierra, lo que se denomina grassroots (base popular). Utilizar esta técnica de manera artificial para hacer campañas sobre la población se consideró como un césped artificial, adoptando el nombre de una marca de dicho producto: astroturf. El astroturfing es, por tanto, una simulación de un movimiento "desde abajo", que procede desde la población en general, cuando en realidad es artificial y creado con unas claras intenciones (García-Estévez, Ballesteros-Aguayo, & Colussi, 2025). Conceptualmente, la técnica explota la teoría de la fortaleza de los enlaces débiles de la teoría sociológica de Granovetter, ya que introduce mensajes a través de personas percibidas como parecidas o cercanas, aunque no mantengan lazos fuertes con la audiencia, para propagar influencia de forma eficiente.

La técnica del astroturfing es hoy una de las herramientas más empleadas en la industria de la desinformación y suele desarrollarse en fases bien definidas (Arce-García, Said-Hung, & Mottareale-Calvanese, 2023). En su primera fase, en un corto periodo varias cuentas que previamente se han conectado con miembros de una comunidad por gustos, barrio, eventos locales o aficiones (el tiempo, el fútbol del pueblo, las fiestas de la ciudad, etc.) publican mensajes muy parecidos, creando la apariencia de espontaneidad. No se recurre a grandes influencers con cientos de miles de seguidores (que ya no se perciben como "iguales"), sino que en su lugar se utilizan nano- y *micro-influencers*, cuentas con cientos o pocos miles de seguidores que comparten afinidades con la audiencia objetivo. Esta distribución dispersa dificulta su identificación con herramientas de análisis de redes, porque las cuentas originales no tienen importancia global en la red. Además, las cuentas que lanzan estas campañas suelen ser gestionadas por personas reales, pagadas y a menudo manejando varias cuentas desde oficinas conocidas como "granjas de trolls", empleando narrativas de inicio de

campaña. En algunos contextos se las llama "usuarios alfa", mientras que otros autores (Keller, & et al., 2019) denominaron a esta primera etapa como "*co-tweeting*" y el grupo de análisis de Propaganda Computacional de la Universidad de Oxford la clasificó como "distribución", etiquetas que apuntan a la misma coordinación sutil para simular apoyo "desde la base".

Tras la difusión inicial, se activa la fase que el equipo de Oxford denomina "amplificación", donde se intenta atraer a grandes influencers, periodistas y medios para que redifundan el mensaje. Se emplean mensajes como "¿habéis visto esto?", "¿qué nos podéis decir de esta información?", etc., para intentar que se fijen en la campaña e incluso se hagan eco de la misma. Si lo logran, el alcance crece con rapidez y la narrativa gana credibilidad, pero si no, recurrirán a portales web que simulan ser medios veraces pero que en realidad difunden desinformación. En esta etapa intervienen trolls o usuarios "beta" (como en la fase inicial) (Bot Ruso, 2019) que responden a cuestionamientos y protegen la campaña ante críticas, empleando narrativas de contrarréplica.

Si una campaña es desmentida en esta fase su impacto es casi nulo, pero debido a su naturaleza y dispersión, muy difícil de identificar. Aún así existen ejemplo como el bulo que difundió que el entonces Ministro de Ciencia de España y ex-astronauta, Pedro Duque, estaba hospitalizado el 21-22 de marzo de 2020 en época de pandemia covid en Denia (Valencia), con mensajes en X/Twitter como "¿Alquien puede verificar esta noticia? ¿Qué hace Buzz Light Year en Denia? ¿No está prohibido el desplazamiento de personas? ¿@astro_duque, donde estás? Pedro Duque ingresa en el Hospital de Denia", en su primera fase, junto con mensajes de segunda fase "@okdiario Pedro Duque (el ministro astronauta) ingresado en el hospital de Denia... se había ido de fin de semana a Jávea!!! Atajo de imbéciles irresponsables. es cierta esta noticia? Investigad". Al aparecer el Ministro rápidamente en televisión, desmintiendo que ni siquiera estaba en Valencia y que estaba sano, la campaña terminó de forma abrupta.

Si las dos primeras fases muestran indicios de éxito, se pasa a la llamada "inundación" (o "*co-retweeting*" según Keller), que suele programarse para coincidir con los picos de uso de redes (desayuno, comida y cena) para maximizar su visibilidad. Cuando la campaña funciona, se repiten oleadas de inundación en horas posteriores o días siguientes hasta concluir la operación, aunque raramente se lanzan más de dos o tres oleadas sin renovar o ajustar el mensaje, ya que el mismo mensaje no suele estar más de 36 a 48 horas activo. En este momento es habitual el uso de bots automatizados para amplificar y sostener la conversación.

La técnica del astroturfing necesita tiempo, ya que las cuentas iniciadoras de la campaña primero deben haberse integrado en los filtros burbuja de otros usuarios, conectar con perfiles de intereses semejantes y haberse ganado cierta confianza previa hablando de otros temas, como el fútbol, música o la televisión. Ese proceso crea una mínima confianza que, además, permite recopilar datos y perfilar mejor a los usuarios que interaccionan con ellas. Una vez consolidada la segmentación de dichos usuarios, se les utilizará dentro de las campañas dirigidas. Lleva tiempo implantarla, pero una vez establecida tiene mucha fuerza y capacidad de influencia, y eso se nota cuando esos nano-influencers empiezan a crecer en seguidores, como empieza a suceder en muchas campañas en España.

Un estudio publicado en 2022 en *Scientific Reports* (Nature) por Schoch et al. (2022) sobre coordinación de campañas a escala global analizó casi 40 procesos electorales entre 2018 y 2020 y encontró que, de media, el 74% de las cuentas involucradas usaban técnicas de astroturfing en Twitter. Los mayores niveles de uso se detectaron, en orden descendente, en Corea del Sur, Honduras y España, con cerca del 90% de las cuentas operando de forma orquestada durante periodos electorales.

Existe una variante del astroturfing llamada "ataque mariposa", en la que cuentas aparentemente inconexas, de baja influencia y con apariencia de "gente normal" se infiltran en una comunidad o burbuja ya existente y establecida para generar en su interior tensiones y mal ambiente. Su objetivo es desestabilizar el grupo y desprestigiar a las cuentas que lideran el discurso.

Cuando no se recurre al astroturfing, la fase inicial de distribución de contenido suele ejecutarla un *influencer* conocido con gran alcance. El desarrollo por fases es parecido, ya que la difusión arrancará desde esa cuenta influyente (periodista, figura pública, etc.) y, si funciona, continuará con las fases siguientes (amplificación con respuestas organizadas, e inundación) siguiendo esquemas y tiempos similares a los del astroturfing. Un esquema muy común es que el influencer principal emita un mensaje original y ya no se involucre más para evitar ser fuente de ataques o contestaciones, siendo 3 ó 4 influencers más pequeños los que continúen el proceso de amplificación y sean los que sirvan de referente y parapeto en las discusiones que surjan.

Un ejemplo ilustrativo ocurre al analizar cuando algún periodista de RT publicó mensajes señalando a varios investigadores españoles. Tras esa publicación inicial, unas cuatro o cinco cuentas continuaban difundiendo el contenido atrayendo la conversación y, en las 48 horas siguientes, varios centenares

de cuentas se sumaban para apoyarlo en fase inundación a horas muy concretas. El seguimiento de esas cuentas durante meses muestra que se empleaban repetidamente para respaldar campañas en distintos países de Latinoamérica.

2.4. LA EJECUCIÓN DE LA CAMPAÑA

Una vez que ya se pasaron por todos los pasos previos, se entra en la fase de ejecución de la campaña, que sigue igualmente varias fases y posibilidades.

2.4.1. Cebado de bomba, entrega de contenido y maximización de la exposición

A la hora de evaluar la viabilidad y posible eficacia de una campaña, resulta esencial analizar su rendimiento previo mediante la difusión controlada de mensajes y narrativas sobre una población de referencia. Esta fase experimental permite comprobar cómo distintas formulaciones discursivas inciden en la percepción y el comportamiento social dentro de un entorno limitado. Para ello, suelen escogerse comunidades bien delimitadas o grupos relativamente homogéneos, en los que sea posible observar con claridad las consecuencias de la intervención comunicativa. En este sentido, las pequeñas localidades (de entre tres mil y cinco mil habitantes) constituyen espacios idóneos para este tipo de ensayos. Su tamaño permite un seguimiento preciso del impacto de la campaña, la definición de métricas de éxito y la evaluación de eventuales estrategias complementarias, como la obtención de financiación o la creación de sitios web específicos. No es casual que proliferen en internet multitud de páginas locales que difunden noticias de municipios y villas españolas ausentes en la agenda de los grandes medios, ya que estos espacios actúan, en muchos casos, como terreno fértil para la experimentación con mensajes dirigidos, adaptados a las sensibilidades o narrativas propias de cada comunidad.

De hecho, en los últimos años se ha especulado con la realización de pruebas que incluso podrían haber incorporado tecnologías de clonación de voz en algunas pequeñas localidades españolas. En ellas, se habría recurrido a imitaciones de voces de personas conocidas en el entorno local con el fin de evaluar el efecto emocional y la credibilidad que generan al enunciar determinados mensajes, lo que abre un inquietante campo de estudio sobre la manipulación perceptiva y la desinformación personalizada. Su salto desde las redes a toda la opinión pública local, algo que ha sido referenciado en algún estudio académico español (Suau, & Puertas-Graell, 2023), es algo que se constata una y otra vez.

Esta fase preliminar, conocida en inglés como *pump priming*, podría traducirse de manera literal como "cebado de bomba". No obstante, en el ámbito hispanohablante suele emplearse una expresión más coloquial y reconocible como "soltar un globo sonda". Sin embargo, el concepto original posee un sentido más sofisticado y estratégico, pues no se trata únicamente de lanzar un mensaje para tantear la reacción pública, sino de preparar el terreno para una acción comunicativa o política de mayor envergadura, calibrando de antemano sus posibles efectos. De esta manera se harían todos los procesos de campaña hasta su evolución final, de cara a depurar su efectividad e impacto.

En este proceso se enmarcan también el intentar captar a *influencers* y medios legítimos en la campaña para amplificar el tamaño de la red y la velocidad de difusión de los comentarios, en el que se calibran los aspectos emocionales, y evitando posibles saturaciones y rechazos. Se crea el entorno, se siembra la semilla y se espera su impacto, llegando incluso a contratar a empresas del ámbito comercial para medir su impacto en la población diana.

Una vez que se depuró la campaña en la prueba inicial (si se realizó), se produce la suelta de la campaña en toda su extensión, en tiempos, formas, canales de distribución, etc. Comienza la ejecución de todo lo planificado hasta ahora: la difusión de contenidos en redes sociales, medios afines, compartir memes, participar y saturar foros de diarios digitales, atraer a influencers y medios legítimos, utilización de expertos falsos, tiempos de difusión, técnicas de difusión, etc., a la par que se realiza contrarréplica de todos aquellos que pongan en cuestión la campaña. En el caso de que los ataques contra la campaña sean intensos y dirigidos, se puede identificar a los principales actores y se les pueden realizar desde ataques con comentarios críticos, acoso, *doxing* (revelar información personal de los mismos), revelar información confidencial o establecer campañas de boicots y cancelación de oponentes.

2.4.2. Acciones offline en el mundo real

Uno de los factores con mayor potencial para amplificar el impacto de una campaña comunicativa es precisamente su traslado del entorno digital al mundo *offline*, es decir, a la esfera tangible y cotidiana de la vida social. Cuando las acciones gestadas a través de internet y las redes sociales logran transformarse en manifestaciones, protestas públicas o incluso enfrentamientos directos, se produce un ciclo de retroalimentación que genera nuevos mensajes, imágenes y vídeos, intensificando de manera exponencial el alcance de la iniciativa y dificultando la contención de sus efectos.

Este fenómeno, aunque cada vez más relevante, ha sido objeto de escasos estudios rigurosos y evidencia empírica concluyente, en parte debido a las sofisticadas técnicas de ocultación y fragmentación empleadas por sus promotores. Sin embargo, existen investigaciones que arrojan luz sobre ejemplos de esta modalidad de manipulación social. Tal es el caso recogido en el informe "*The Tactics & Tropes of the Internet Research Agency*", elaborado en 2019 por un equipo de expertos a propuesta del Comité de Inteligencia del Senado de los Estados Unidos (DiResta, & et al., 2019). En dicho documento se describen episodios en que los agentes del *Internet Research Agency* (IRA) organizaron protestas opuestas en un mismo lugar (como ocurrió en Nueva York) llegando incluso a contratar ciudadanos locales para que se encargaran de la organización y participación en dichos eventos, lo que derivó en enfrentamientos directos entre los asistentes.

El mecanismo de acción revelaba una planificación meticulosa, donde en un primer paso se creaban perfiles y páginas falsas en redes sociales, tanto para colectivos afines a "*Black Lives Matter*" como para agrupaciones contrarias (como el movimiento "*Blue Lives Matter*"). Estas estructuras virtuales facilitaban la convocatoria simultánea de eventos rivales en espacios públicos, seleccionando fechas y localizaciones coincidentes, o fomentando divisiones y polarización. Entre los casos más ilustrativos se encuentra el enfrentamiento frente a una mezquita en Houston en 2016, originado por la acción coordinada de perfiles falsos gestionados por el *Internet Research Agency*. En esa ocasión, se orquestaron protestas simultáneas de grupos opuestos a través de Facebook y Twitter, derivando en un choque presencial intencionado que más tarde fue amplificado mediante la difusión masiva de imágenes y vídeos en redes sociales (Riedl, et al., 2021).

Asimismo, distintos medios europeos como *Le Monde* (2023), *Süddeutsche Zeitung* (Traufetter, Hesse, & Jung, 2023) y *Deutsche Welle* (2023) han revelado que entre 2022 y 2023 agentes vinculados al Kremlin han simulado manifestaciones e infiltrado en protestas reales en capitales como París, Bruselas, Frankfurt, Berlín, Madrid y otras ciudades relevantes. A través de ello se han identificado expresiones de polarización anti-Ucrania y anti-Turquía en distintas marchas, con la presencia de individuos que realizaban saludos nazis y portaban banderas ucranianas, acciones orientadas a generar confusión y conflicto en la opinión pública local. Los reportajes constatan que los mismos provocadores aparecieron posteriormente en manifestaciones de diferente índole, como las protestas contra la reforma de pensiones en Francia, donde llevaban pancartas exigiendo el cese del apoyo occidental a la guerra en Ucrania. Este

fenómeno demuestra la capacidad de estas redes para apropiarse de causas sociales diversas y reorientarlas según sus intereses, amplificando divisiones y proyectando narrativas afines a sus objetivos estratégicos.

El objetivo no es otro que promover la presencialidad física, intensificando el conflicto en la calle como parte de una estrategia deliberada para "sembrar división literal" (*sow literal division*). De este modo, cuestiones como los enfrentamientos raciales, los movimientos secesionistas o las protestas entre partidos políticos se trasladan del espacio digital al entorno material, generando disturbios cuya documentación audiovisual (vídeos, fotografías y mensajes) alimenta sucesivas fases de la campaña y refuerza su alcance social. El ejemplo del inicio de este libro sobre las vacaciones de Pablo Iglesias e Irene Montero, en el que una frutería en Langreo (Asturias) sirvió de desvío de atención y extensión de la campaña, es igualmente un indicador de que las acciones *offline* están a la orden del día. También en España se pueden encontrar acoso a personal de la Agencia Española de Meteorología (AEMET) (Núñez, 2025), al igual que ha sucedido de la misma manera con meteorólogos de otros países como Francia, Estados Unidos, Alemania o Portugal (Maldita, 2024).

Para explicar este fenómeno podemos considerar que entra en juego la teoría de Paul Brass, un politólogo de EEUU de la Universidad de Washington, sobre los "sistemas institucionalizados de disturbios" (MacMillan Center Yale, 2022). Brass sostenía que los episodios de violencia colectiva no eran erupciones espontáneas de ira popular, sino el resultado de un proceso cuidadosamente orquestado, en el cual ciertos actores políticos (los llamados *fire tenders* o creadores del fuego, y los *conversion specialists* o difusores de dicho fuego) mantienen viva la tensión social y la identidad enfrentada de los grupos. Lejos de limitarse a simples agitadores, estos individuos y organismos funcionan como verdaderos gestores de la violencia, guardianes de una combustión latente que puede ser encendida cuando el contexto lo favorece o los intereses lo exigen. En los estudios de campo que Brass desarrolló durante décadas en la India, especialmente en Uttar Pradesh, observó que los rumores eran la materia prima del conflicto, pues eran instrumentos flexibles, capaces de transformar incidentes triviales en símbolos de humillación colectiva o de amenaza existencial, listos para ser amplificados por quienes dominan la narrativa pública.

Brass clasificaba en tres fases este tipo de sucesos (Grazda, 2024): (1) en la fase preparatoria, los actores de la tensión mantienen un flujo continuo de estímulos emocionales y simbólicos que alimentan la desconfianza; (2) en la activación, el incidente funciona como catalizador que desencadena la moviliza-

ción, ahora amplificada por algoritmos y coordinada mediante redes de mensajería instantánea; (3) y finalmente, en la fase de interpretación, los propios instigadores moldean la memoria del evento, redistribuyendo culpas y reescribiendo los hechos para perpetuar el ciclo.

Con la llegada de las plataformas digitales, los *fire tenders* no necesitan presencia física, pues pueden ser figuras con legitimidad simbólica o incluso identidades difusas tras perfiles anónimos, que alimentan el miedo, la desconfianza o la indignación desde redes sociales o canales de mensajería cifrada. Su tarea, ahora multiplicada por la lógica del algoritmo, consiste en mantener viva la polarización cotidiana, construyendo identidades defensivas y reforzando la sensación de agravio constante. Los *conversion specialists*, por su parte, actúan cuando una chispa (una noticia manipulada, un video viral o una acusación moral) se convierte en oportunidad. Son quienes redefinen el hecho, lo dramatizan y lo elevan al rango de causa colectiva, son los que logran que una emoción digital se traduzca en cuerpo callejero, en masa convocada y, a menudo, en violencia física.

Los efectos *offline* de este proceso, observables en contextos tan distintos como los disturbios de Delhi (India) de 2020, el asalto en Washington D.C. al Capitolio estadounidense en 2021, o graves altercados por inmigración en Southport, Inglaterra, en agosto de 2024 y en Torre Pacheco en Almería, España, en julio de 2025 (con multitud de bulos a través de canales de ultraderecha y de la red *Pravda* pro-Kremlin con un supuesto comunicado del Ayuntamiento que decía que la inmigración provocaba inseguridad y otros bajo el mensaje "Solo el pueblo salva al pueblo") (Sanchis, & Alamillos, 2025) confirman que las fronteras entre el rumor digital y el acto violento físico y real se han difuminado.

Las redes funcionan como incubadoras de pertenencias radicalizadas, donde el "otro" deja de ser un adversario discursivo y se convierte en una amenaza ontológica. Así, la desinformación no solo deforma percepciones, sino que a través del odio produce condiciones materiales para la agresión, ya que justifica, moviliza y legitima la acción colectiva violenta. El mecanismo que Brass identificó en los años noventa del siglo XX, articulado en fases de preparación, activación y reinterpretación, se mantiene vigente, aunque sus escenarios de actuación han migrado al espacio virtual, de las asambleas políticas a las burbujas informativas. Y el salto desde el mundo virtual al real, con el consecuente aumento de delitos por crimen de odio en el mundo real ya está demostrado en España, al comparar los niveles de odio en Facebook con los registros policiales de 2016 a 2018 (Arcila, & et al., 2024).

2.4.3. Monetización

Las estrategias de monetización en la industria de la desinformación destacan por su capacidad de adaptación y su extraordinaria diversidad. Se despliegan como un entramado de tácticas técnicas y procedimentales orientadas a maximizar beneficios, ocultar el origen de los fondos y mantener una apariencia de legitimidad. Estas prácticas abarcan desde el uso de mecanismos tradicionales (como la publicidad o la venta de productos) hasta métodos más opacos y tecnológicamente sofisticados, entre ellos el empleo de criptomonedas, la recaudación mediante plataformas de micromecenazgo y la comercialización de datos personales.

El equilibrio entre el bajo coste y la alta rentabilidad constituye uno de los pilares de la economía de la desinformación. La generación de narrativas manipuladas exige una inversión mínima pues no requiere contrastar fuentes, verificar datos ni sostener procesos de investigación complejos. Sin embargo, su difusión puede alcanzar proporciones masivas y generar beneficios significativos a través del tráfico digital y la publicidad asociada, ya que de este modo, la desinformación se configura como un negocio de inversión reducida y retorno elevado, sostenido por las métricas de interacción que alimentan los algoritmos de las plataformas digitales. A ello se suma el anonimato como elemento funcional del sistema, pues protege a los emisores y dificulta cualquier intento de trazabilidad. Las operaciones mediante plataformas *offshore* y criptomonedas refuerzan este entramado opaco, y aunque los avances en el análisis de *blockchain* han permitido identificar ciertas rutas financieras, la detección plena de los actores implicados sigue siendo todavía un desafío considerable.

Entre las estrategias más frecuentes, analizados en el tercer capítulo del informe del Foro de Lucha contra la Desinformación de 2024 (Galan Cordero, & et al., 2024), están los pagos en criptomonedas y la financiación a través de plataformas de *crowdfunding*. Los actores de la desinformación recurren a carteras digitales y servicios como *Patreon*, *Tipee* o *GoFundMe*, favoreciendo un flujo de donaciones difícil de rastrear que sostiene tanto la infraestructura tecnológica como campañas específicas. Esta capa de anonimato financiero les permite operar con aparente autonomía y escasa exposición legal.

La publicidad en sitios web y redes sociales constituye otro pilar sustancial del modelo económico desinformativo. La creación de portales temáticos (ya sean supuestamente informativos, de entretenimiento o de contenido político) atrae tráfico masivo y, con ello, anunciantes dispuestos a pagar por un público amplio y barato. Se estima que el 70 % de estas páginas difunde publicidad de

negocios y alrededor del 40 % promociona contenido de ocio. Muchas plataformas al priorizar la viralidad en sus algoritmos, contribuyen de manera indirecta a amplificar los mensajes sensacionalistas que más ingresos generan. A esta dinámica se suma el mercado clandestino de compraventa de cuentas robadas, especialmente activo en la web oscura o *darknet*. Estas identidades sustraídas permiten amplificar mensajes, manipular conversaciones y suplantar perfiles auténticos, expandiendo el alcance y la credibilidad de las campañas. En paralelo, empresas de consultoría y *think tanks* desempeñan un papel discreto pero decisivo, ya que canalizan recursos mediante entidades pantalla o estructuras financieras opacas que dificultan la trazabilidad de los fondos y su vinculación con los verdaderos promotores.

La mercantilización de la propaganda alcanza también el terreno simbólico. Muchos de estos actores comercializan merchandising, libros o artículos que refuerzan su narrativa ideológica y consolidan su comunidad de seguidores, transformando la desinformación en una marca rentable. Igualmente, la recopilación y posterior venta de datos personales (obtenidos a través de interacciones digitales o formularios manipuladores que han definido las segmentaciones iniciales para preparar las campañas) representa una fuente adicional de ingresos, al alimentar el mercado de información y segmentación publicitaria. Un ejemplo lo describe el informe sobre las operaciones del IRA para el Senado de EEUU, en el que se describe cómo se vendían datos de usuarios y de segmentación destinados a empresas de marketing, así como productos falsos y merchandising (venta de productos como cuadros, pegatinas, carteles o camisetas ideológicas y nacionalistas) que generan comunidades afines y hacen aumentar su credibilidad y autenticidad (DiResta, 2019).

Hay muchos otros ejemplos, desde 2020 a 2022 el *Children's Health Defense* (CHD) recaudó 46 millones de dólares realizando campañas con multitud de bulos y consignas antivacunas como "defensa de la libertad" y "lucha contra la censura", mientras que otras, con las mismas consignas como *Informed Consent Action Network* (ICAN) llegaron a 13,4 millones (El Economista, 2024). Canales de Youtube que promueven teorías conspiranoicas como Parafantástico, que habló sobre la teoría del microchip y la vacuna del Covid, tenía ganancias estimadas en 692 a 11.100 dólares mensuales, mientras que el canal *DrossRotzank* con diversas teorías "alternativas" se estima pueda llegar de 18.300 a 293.100 dolares al mes (Verificado, 2020).

Durante la pandemia, se registró el caso de una doctora española conocida por su postura negacionista frente a la existencia de Covid-19 y por ser

una de las fundadoras del colectivo "Médicos por la verdad". Esta organización, que abanderaba discursos contrarios al consenso científico, logró recaudar más de 4.000 euros en donaciones con el fin de financiar campañas de desinformación y tratar de expandir su influencia hacia América Latina (Equipo de Investigación, 2021). En Telegram, el grupo aprovechó y reutilizó el canal "Rafapal Amigos", que era ampliamente reconocido en los entornos conspiranoicos debido a su vinculación con Rafapal, un creador de contenido que se autodefine como practicante de "periodismo para mentes cósmicas" y que destaca por la difusión sistemática de teorías conspirativas originadas en Estados Unidos. Además, la estrategia de difusión y captación de nuevos seguidores incluyó la diversificación de los canales de comunicación hacia plataformas como *Rumble*, *Odysee* y *Bitchute*, espacios digitales que han adquirido notoriedad por albergar y facilitar la propagación de campañas conspiranoicas y discursos alternativos (Madrigal, 2021).

Otro buen ejemplo se puede ver con la investigación de la empresa española Newtral sobre la web "Invadidos.com" (Newtral, 2025), que evidencia cómo esta web antiinmigración materializa los modelos de negocio opacos y rentables propios de la industria de la desinformación, empleando identidades ficticias y proxies ("Los proxies o actores interpuestos son entidades, organizaciones o individuos dentro de un Estado que actúan en interés de un actor extranjero. Pueden actuar como supuestos medios de comunicación, empresas de márketing y publicidad, organizaciones políticas, grupos de interés, funcionarios, o incluso figuras públicas e *influencers*") (Arce-García, & et al., 2024) legales para ocultar a los responsables reales, financiándose mediante campañas de donaciones y transacciones con criptomonedas, mientras maximiza sus ingresos gracias a un alto volumen de publicaciones y viralización en redes sociales, amplificando narrativas de odio y contribuyendo a la polarización social.

2.4.4. Lavado de información

El lavado de información constituye una estrategia principal dentro del repertorio de técnicas de desinformación contemporáneas, cuyo propósito es legitimar y propagar contenidos manipulativos mediante una compleja red de intermediarios informativos. Este procedimiento está hecho para enmascarar el origen de la información, facilitando su entrada en el discurso público y dificultando su rastreo y desarticulación por parte de las autoridades y los medios legítimos, y generalmente suele estructurarse en tres etapas diferenciadas (Stolze, 2022):

- Emplazamiento inicial: En la primera fase, los contenidos manipulativos se introducen en canales de comunicación afines, poco vigilados o periféricos, como foros digitales, blogs o medios alternativos, donde pueden diseminarse sin restricciones ni controles significativos.
- Superposición o amplificación intermediaria: Una red de intermediarios interconectados (compuesta por plataformas, perfiles y otras páginas informativas) se encarga de republicar y reinterpretar el mensaje, omitiendo deliberadamente cualquier referencia a la fuente original. Este paso es crucial para "blanquear" el contenido, dificultando la atribución y dotando al mensaje de mayor legitimidad aparente.
- Integración pública: Finalmente, el mensaje transformado es recogido y amplificado por medios convencionales o por actores reputados, quienes, aunque en ocasiones lo mencionen fuera de contexto (efecto "woozle"), contribuyen de manera decisiva a su legitimación y a su difusión masiva entre el público general.

La literatura reciente y los informes especializados de organismos como StratComCOE (Servicio de Comunicaciones Estratégicas de la OTAN) (Carrasco Rodríguez, 2023) destacan una batería de tácticas adicionales empleadas en campañas de lavado de información, como son el uso de dominios nacionales considerados de confianza, para conferir credibilidad al mensaje, la apropiación indebida (*misappropriation*) de datos reales, mezclados con información falsa, con el fin de dificultar la detección de la manipulación y, por último, la suplantación de identidades o entidades mediáticas (*impersonation*), utilizada para conferir autoridad y verosimilitud a los contenidos manipulados.

En estudios de caso como en Alemania (Carrasco Rodríguez, 2021), se ha observado cómo actores vinculados al Kremlin utilizan medios locales (aprovechando su arraigo y conocimiento del entorno informativo) para insertar narrativas distorsionadas. El patrón recurrente consiste en la reutilización de noticias auténticas, extraídas de medios alemanes populares, que son posteriormente modificadas y enriquecidas con elementos engañosos en medios pro-Kremlin, con el fin de influir y manipular la opinión pública alemana. También se reportaron casos similares en Suecia (Pammet, & et al., 2019) con las mismas técnicas en torno a la isla de Gotland, donde medios rusos como *Sputnik* publicaron en 2015 y en varios idiomas, noticias manipuladas que distorsionaban declaraciones oficiales suecas y militares, usando fuentes originales pero eliminando el contexto y alterando su significado para crear alarma sobre un ataque sueco con misiles a Rusia.

Un fenómeno lateral del lavado de información es el de las operaciones de blanqueo reputacional, para el borrado de noticias e información que encubren, desvíen o legitimen información perjudicial. De esta manera, existen empresas cuya misión es eliminar, ocultar o alterar resultados de búsqueda y huella digital de clientes con antecedentes adversos, incluyendo políticos corruptos, empresarios y organizaciones criminales. Un buen ejemplo se tiene en la empresa española Eliminalia, actualmente iData Protection, que una agrupación de periodistas llamada *Forbidden Stories* pertenecientes a varios diarios de todo el mundo (entre ellos de El País, *The Washington Post*, *The Guardian*, *Der Spiegel* y *Haaretz*) (OCCRP, 2023) identificó cómo empleaba noticias falsas, webs clonadas y bots para engañar a webs como Google, o desaparecer de la Wikipedia (A.G.P., 2023). Se conoce que gente de 54 países, entre ellos gobernadores mexicanos, narcos y jerarcas chavistas ficharon a esta empresa de desinformación para lavar su imagen en la red (Gil, & Irujo, 2023a). Filtraciones de reportajes de 2023 revelaron más de 1.500 clientes, de los cuales más de 700 eran procedentes de España, incluyendo a un sacerdote arrestado por saquear a viudas, un policía condenado por violencia, empresarios salpicados en escándalos de corrupción o un exmiembro del Consejo General del Poder Judicial (CGPJ) (Gil, & Irujo, 2023b).

2.5. EVALUACIÓN DE LA EFICACIA DE LA CAMPAÑA

Una vez concluida la fase de difusión de una campaña, resulta imprescindible evaluar con rigor si esta ha alcanzado los objetivos inicialmente planteados. Para ello es fundamental elaborar un proceso de medición sistemático mediante el uso combinado de métricas cuantitativas y cualitativas, que permitan determinar de forma objetiva tanto el grado de consecución de las metas propuestas a lo largo de la campaña como la existencia de posibles desviaciones. Este análisis no sólo contribuye a identificar los aciertos y áreas de mejora al finalizar la acción, sino que proporciona aprendizajes valiosos para la optimización de futuras estrategias. En este contexto cobran especial relevancia los denominados KPI (*Key Performance Indicators*, o indicadores clave de rendimiento), cuya formulación debe estar alineada con los objetivos iniciales, garantizando que sean alcanzables, medibles, relevantes, acotados temporalmente y bien definidos. Dichos indicadores han de monitorizarse tanto durante el desarrollo de la campaña como al término de la misma, permitiendo un seguimiento continuo y una evaluación final exhaustiva.

Este proceso de evaluación requiere la definición de marcadores específicos que reflejen el desempeño real de la campaña, donde se pueden destacar el

alcance conseguido, la identificación de los grupos destinatarios, el grado de engagement o compromiso logrado entre los seguidores, así como la naturaleza y cuantía de las reacciones, haciendo especial hincapié en la dimensión emocional de las mismas. Asimismo, resulta esencial analizar si la campaña ha conseguido penetrar en colectivos clave, como influencers o medios de comunicación de prestigio, capaces de amplificar su impacto. Un indicador relevante es la eficacia de las técnicas empleadas previamente para el conocimiento y segmentación del público objetivo a través de la teoría de redes, que permite medir, entre otras cosas, el grado de egocentrismo en la difusión (es decir, si el contenido ha saltado a nuevos grupos sociales), identificar las cuentas que han actuado como intermediarias y aquellas que han liderado la conversación. Todo ello facilita, de cara a futuras campañas, la detección de nuevos usuarios de los cuales pueden inferirse patrones de gustos y tendencias a partir de sus interacciones, ya sea conversando o viendo sus reacciones posteriormente sobre fútbol, música u otros temas de interés. Estos perfiles pueden ser estudiados con mayor profundidad, e incluso llegar a convertirse en activos ignorantes o integrantes de cámaras de eco para aprovecharlas en nuevas estrategias comunicativas.

Otro aspecto fundamental a considerar es el universo emocional y narrativo que se desprende de las cuentas que participan en la campaña, tanto desde posiciones favorables como críticas. El análisis detallado de las emociones y relatos generados en estas interacciones se revela clave para la construcción y posterior refinamiento de nuevas estrategias comunicativas, permitiendo ajustar las narrativas en futuras campañas con mayor precisión. Para ello, es esencial definir previamente los indicadores clave de rendimiento (KPI) correspondientes, que orientarán la observación de diversos ámbitos.

La identificación de cambios en el comportamiento de los usuarios resulta central, ya que una mayor carga emocional y la presencia de nuevos discursos pueden indicar una creciente susceptibilidad a la introducción de ideas novedosas. Este seguimiento posibilita advertir modificaciones en el conocimiento, la conciencia colectiva y las respuestas ante determinados temas dentro de la población objetivo:

- Identificación de usuarios y grupos que tengan una mayor influencia sobre su entorno, o que sean más susceptibles ante los efectos de la campaña.
- El tiempo en el que empiezan a observarse modificaciones en los discursos, variaciones en la polaridad de los mensajes y fluctuaciones emo-

cionales en las respuestas, se convierte en un indicador clave que permite ajustar los tiempos de intervención y perfeccionar el impacto de las acciones en futuras iniciativas. A través de esta observación minuciosa del ritmo de cambio será posible afinar los tiempos de intervención de cara a optimizar el impacto de acciones futuras.

- En lo relativo al alcance, es pertinente analizar tanto la duración como la distancia media y máxima recorrida por los flujos de comunicación, es decir, los saltos que los mensajes atraviesan entre usuarios. Cabe recordar que Stanley Milgram, en 1967, postuló la existencia de seis grados de separación entre cualquier par de individuos en el mundo (Milgram, 1967), concepto que más tarde fue validado en varios estudios realizados en redes sociales digitales, como Facebook (con valores de 4,74 en 2011 y 3,57 en 2016) y X/Twitter (4,67) (National Geographic España, 2022). Una evaluación precisa de estas distancias máximas proporciona indicios claros sobre el nivel de cohesión de la red de usuarios alcanzada por la campaña.
- Cambios que trascienden el ámbito digital, ya que se observarán los cambios y hechos relevantes que saltaron desde el mundo online al mundo real, en un proceso de transferencia de la campaña de desinformación y/o odio hacia la sociedad.

Una vez verificada la eficacia y el grado de cumplimiento de los objetivos, es recomendable revisar y actualizar las narrativas empleadas, las estrategias de difusión, los marcos temporales y los prejuicios explorados, con el propósito de lograr un impacto aún mayor y perfeccionar las futuras campañas.

2.6. TECNOLOGÍA DE DISTRIBUCIÓN DE MENSAJES

Las compañías especializadas en la gestión de granjas de trolls y bots operan apoyándose en una sofisticada infraestructura física y digital, concebida para administrar miles de cuentas automatizadas en redes sociales. Su objetivo es doble, ya que por un lado pretenden maximizar la eficiencia operativa, y por otro salvaguardar el anonimato de las actividades, evitando la detección por parte de las plataformas.

En el plano físico, destacaba en una primera época el uso de paneles llenos de móviles que eran controlados por una persona, para pasar a una segunda época en la que se pasaron a paneles equipados con decenas de teléfonos conectados de forma simultánea, cada uno operando múltiples cuentas y operados

desde un ordenador. En una tercera generación, los móviles se insertan ya en racks o cajas-contenedores donde los equipos contienen las placas de los teléfonos móviles, ya sin carcasa y ni siquiera pantalla, conectados de manera que se maximice su espacio y su funcionamiento con control de temperatura incorporada. Estos contenedores suelen venderse ya preparados para contener cada uno en torno a 20 teléfonos, con un procesador informático propio controlador para cada uno con un costo entre 150 a 1.500 € (precio de 2025) de cada conjunto dependiendo de la calidad de los equipos empleados (almacenamiento, RAM, número de procesadores, etc.) (Cellhasher, 2025).

Figura 13. De izquierda a derecha, granjas de primera a tercera generación.
Fuente: @DouglasMun (2024)

La manipulación directa de estos dispositivos permite acciones como el intercambio rápido de tarjetas SIM, la restauración de fábrica y otros trucos destinados a dificultar la traza y el rastreo de las cuentas. La gestión de tarjetas SIM, provenientes de diversas operadoras, se combina con el empleo de routers WiFi específicos que simulan ubicaciones, multiplicando las procedencias aparentes y sorteando bloqueos basados en patrones de acceso. Complementando este sistema, los servidores *proxy* y las redes VPN se integran para camuflar aún más la ubicación real tanto de los dispositivos como de las cuentas, permitiendo a las empresas operar de manera simultánea desde múltiples puntos virtuales. De esta manera pueden controlar incluso la posición geográfica GPS o la dirección IP informática para hacer creer que se encuentran en el lugar y país que consideren conveniente.

La cuarta generación se está implantando con granjas virtuales en la nube, gestionadas generalmente por empresas del sudeste asiático, y ofrecen sus ser-

vicios de móviles virtuales con especificaciones para operar en redes como Facebook, Tiktok, Instagram o Spotify, mientras que los clientes operan sólo, de forma remota, el gestor final de contenidos desde su ordenador. Estos sistemas en la nube ofrecen el uso de 1.000 teléfonos móviles con cuentas en una red social por un coste de unos 27$ al mes, o 250$ al año (Genfarmer, 2025).

El coste de tener cuenta en cada red social requiere confirmación de un SMS a un número de teléfono válido, por lo que hay un mercado de venta de cuentas ya verificadas para operar en cada red o plataforma y en cada país. Según investigadores de la Universidad de Cambridge y su *Cambridge Online Trust and Safety Index*, el coste de una cuenta ya verificada sería muy distinto según día, hora y lugar. Así, cada cuenta verificada costaría de media el 14 de diciembre de 2025 0,03$ en Indonesia, 0,29$ en Rusia, 0,04$ en Reino Unido, 0,18$ en Estados Unidos, 0,26$ en China, 0,10$ en España, 0,03$ en Australia o 1,45$ en Japón (aunque los precios varían mucho en el tiempo). En concreto, en esa misma fecha había disponibles en el mercado negro un total de 2.559.937 cuentas ya verificadas listas para comprar y usar en el ámbito del mercado español, a un precio de 7$ por una cuenta de Whatsapp, 3,23$ por una de Telegram, 2,12$ una de Bet365, 1,79$ de Paypal, 0,81$ de Tinder, 0,76$ de Idealista, 0,75$ de Glovo, 0,53$ de Shein, 0,30$ de Facebook, 0,29$ de Google/Gmail/Youtube, 0,13$ de X, 0,11$ de Tiktok o 0,02$ de Linkedin, a modo de muestra. La procedencia de venta de dichas cuentas se efectuaría principalmente desde China, Rusia, Estados Unidos, Indonesia, Nigeria, Bielorrusia y Vietnam (Lewsey, 2025, University of Cambridge Social Decision-Making Lab, 2025).

El *software* de automatización multisistema posibilita la gestión simultánea de miles de interacciones diarias (*likes* o me gusta, comentarios, seguimientos, visualizaciones) segmentadas por cuenta, red social y campaña concreta. Los bots integran algoritmos de inteligencia artificial y *machine learning* capaces de analizar tendencias, generar contenidos adaptados y simular tanto el lenguaje como los patrones de socialización humana, lo que complica notablemente su identificación. Herramientas como *crawlers* y *scrapers* operan a gran escala, rastreando información social para identificar objetivos, palabras clave, narrativas emergentes y vulnerabilidades en las plataformas. Todo esto se controla desde paneles de gestión centralizados que permiten programar tareas, monitorizar el rendimiento de las cuentas, analizar el retorno de inversión y coordinar respuestas automáticas y ataques dirigidos.

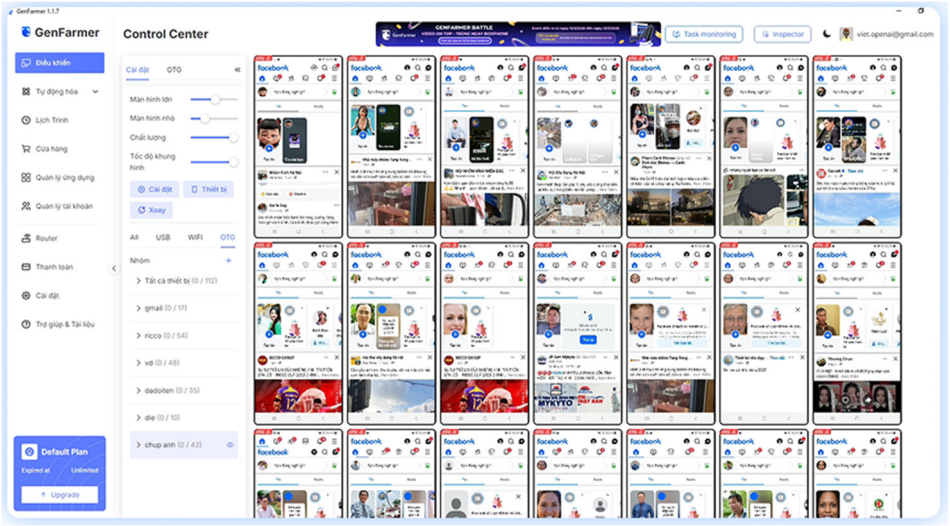

Figura 14. Programa de gestión de teléfonos virtuales en alquiler en nube, desde una aplicación de control. Fuente: Genfarmer (2025)

Estas empresas desarrollan incluso aplicaciones propias, como software complementario para dispositivos Android e iOS, optimizadas para la manipulación social y la evasión de filtros y protecciones de las plataformas. Estos programas emplean sistemas que simulan el comportamiento humano real, imitando la navegación pausada, la lectura, la reproducción de vídeos y el movimiento de pantalla, lo que incrementa la verosimilitud de los bots ante controles automatizados. Por si fuera poco, estas empresas ofrecen incluso capacitación y soporte técnico personalizado, impartiendo cursos, asesoría directa y mantenimiento constante, con el objetivo de potenciar el impacto de las campañas y eludir cualquier intento de bloqueo. Un ejemplo ilustrativo de este tipo de empresas, que ofrecen sus servicios para hacer campañas, mejorar resultados en videojuegos o hacer estafas, es la operación llevada a cabo por Europol en colaboración con las policías de Austria, Estonia y Letonia el 10 de octubre de 2025. En éste último país fueron detenidas siete personas que gestionaban cinco servidores y tenían en su posesión 1.200 cajas con 40.000 tarjetas SIM de teléfono activas, además de varios cientos de miles más listas para activar. Este sistema permitía a redes criminales de más de 80 países alquilar más de 49 millones de cuentas en línea (Europol, 2025).

2.7. LA PREVENCIÓN

2.7.1. Pensamiento crítico y alfabetización mediática

Las redes sociales y el internet han cambiado mucho la manera en que nos informamos y tomamos decisiones. Cada vez tenemos más acceso a datos y opiniones, pero eso no significa que entendamos de forma automática qué es verdad y qué no. Mucha gente, especialmente los jóvenes, sabe que existen noticias falsas o retos peligrosos circulando en las redes, pero aun así les cuesta identificar cuándo se enfrentan a información falsa y cómo reaccionar ante ella (Pérez García, & et al., 2024), más aún ante las nuevas tecnologías y las redes sociales (de Gregorio Vicente, 2023).

El problema no es solo la cantidad de noticias falsas o engaños que circulan, sino también cómo funcionan las propias redes sociales, pues los algoritmos nos muestran lo que "queremos ver", reforzando nuestras ideas y emociones, y a menudo nos impulsan a reaccionar rápido, sin pensar demasiado. Por eso, ahora más que nunca la alfabetización mediática y el pensamiento crítico son muy importantes como guías para aprender a filtrar información y preguntarse de dónde viene, si es fiable y cómo afecta lo que creemos. Sin embargo, pocos docentes cuentan con la preparación necesaria para enseñar este tema correctamente, y muchas veces falta tiempo y recursos para hacerlo (Bolaños Noguera, Zambrano Hinestrosa, & Tumal Chaves, 2024).

Es importante practicar analizar cómo funcionan las redes, aprender a crear contenidos que respondan a las mentiras y, sobre todo, reflexionar sobre cómo y por qué tomamos ciertas decisiones al informarnos. Así, el pensamiento crítico en internet y redes se convierte en una herramienta esencial para defendernos de la desinformación, ayudándonos a construir una sociedad más informada y menos manipulable (Gozálvez-Pérez, Valero-Moya, González-Martín, 2022).

La investigación actual demuestra que la docencia centrada únicamente en la alfabetización mediática resulta insuficiente, pues la educación en muchas ocasiones solo enseña qué es una noticia falsa o cómo buscar en Google y poco más. Y esto sucede porque la desinformación no es un problema de información, sino de arquitectura digital y de estructuras de poder.

Estudios realizados en varios países europeos revelan que, aunque los educadores identifican la falta de alfabetización mediática como una brecha crítica, también advierten que las herramientas digitales de verificación profesional necesitan ser reconfiguradas para adaptarse a la complejidad del tema, lo que sugiere que la pedagogía por sí sola no puede resolver un problema diseñado

para maximizar su efecto mediante sesgos cognitivos (Nygren, & et al., 2022). Investigaciones sobre estrategias contra la desinformación en Filipinas, por ejemplo, muestran igualmente que aunque los educadores implementan estrategias a nivel institucional y de aula, las inconsistencias curriculares, la falta de formación profesional específica y los recursos limitados impiden cultivar competencias críticas efectivas, evidenciando que la pedagogía necesita acompañamiento de políticas públicas y reformas estructurales (Baja, & Borja II, 2025). Esto se corrobora con la evidencia de que, durante la pandemia, la proliferación de teorías conspirativas no respondió a un déficit de conocimiento sino a la incapacidad de los procesos metacognitivos para filtrar la inundación emocional y la arquitectura algorítmica que premia la viralidad sobre la veracidad (Suárez-Ruiz, & González Galli, 2022).

Por todo ello, la desinformación no se combate solo enseñando a identificar noticias falsas, sino desarrollando una epistemología estratégica hacia los medios que critique las propias estructuras que permiten la manipulación informativa. Para ello es muy acertado acercarse a lo propuesto por Grohmann y Ong (2024), donde indican que combatir la desinformación implica actuar mucho más allá del *fact-checking* puntual y de la mera regulación de contenidos en plataformas, interviniendo sobre las estructuras económicas, laborales y políticas que sostienen la "desinformación por encargo". Este enfoque propone una estrategia global y que afecte a toda la sociedad, que incluya no solo a las tecnológicas, sino también a sectores como la publicidad, las relaciones públicas, la consultoría política, los mercados de seguidores falsos y toda la cadena de trabajo digital precario que produce y amplifica contenidos manipuladores. Existe un determinismo en focalizar solo en lo malas que son las plataformas y en el bulo en sí, pero centrarse solo en moderación y verificación deja intactas a las organizaciones y modelos de negocio que convierten la desinformación en una industria rentable, y además abre la puerta a que gobiernos autoritarios capturen esos mecanismos.

Entre las herramientas propuestas por estos autores destacan la reforma de la financiación de campañas y la transparencia en el gasto electoral, la monitorización sistemática de bibliotecas de anuncios políticos y la presión coordinada sobre anunciantes para desincentivar la financiación de ecosistemas desinformativos. También se plantean códigos de conducta y modelos co-regulatorios para agencias de propaganda y consultoras políticas, así como obligaciones de trazabilidad sobre quién paga, diseña y distribuye campañas de influencia. En paralelo, se aboga por políticas laborales y tecnológicas que reduzcan la precariedad que empuja a muchos jóvenes a aceptar trabajos de troleo,

gestor de bots o productor de contenido de narrativas tóxicas, fortaleciendo sindicatos, movimientos sociales y mecanismos de protección a denunciantes que revelen estas prácticas.

En el plano comunicativo, este enfoque incorpora estrategias como el "silencio estratégico", que consiste en no responder de forma automática a todos los mensajes extremistas o conspirativos, para evitar amplificarlos y premiar con atención a los actores que operan mediante la lógica de la provocación constante. Esto deberá ir combinado junto con intervenciones más selectivas y contextuales, que prioricen desmontar campañas con gran capacidad de daño sin convertir cada bulo marginal en un tema central de la agenda pública.

2.7.2. Herramientas de verificación y fact-checking

La verificación periodística o *fact-checking*, según García-Marín (2024), se realiza en un proceso de tres momentos encadenados: primero, los equipos se enfrentan a un flujo de bulos y contenidos dudosos y tienen que decidir qué merece ser comprobado, priorizando los mensajes con mayor potencial de daño o impacto social y aplicando criterios de relevancia pública similares a los de las noticias. A partir de esa selección se entra en una fase metodológica en la que cada afirmación se trata casi como una hipótesis científica, donde se formula la sospecha de falsedad, se buscan y triangulan fuentes documentales, testimoniales y expertas, se emplean herramientas digitales como búsquedas inversas o análisis de metadatos y se concluye con un veredicto que permiten distinguir entre lo verdadero, lo engañoso, lo impreciso y lo no comprobable. El resultado se articula en un texto donde se reconstruye de forma transparente el camino seguido por los periodistas para desmontar la desinformación y explicar qué se comprobó, cómo se hizo y por qué se llega a ese juicio concreto. Así, el *fact-checking* no solo desmiente datos falsos, sino que intenta enseñar al lector a pensar como un verificador.

Este proceso también tiene su momento de avance, automatización y uso de las nuevas tecnologías, donde el *blockchain* y las APIs de *fact-checking* (verificación de hechos) abren un nuevo horizonte de posibilidades. Las técnicas de *blockchain* (tecnología de almacenamiento digital distribuido e inmutable, en la que la información se agrupa en bloques enlazados de forma segura en distintos lugares) se basan en aportar trazabilidad y resistencia a la manipulación al registro de noticias y contenidos, creando certificados digitales que autentifican el origen y la autoría de información (Coface, 2022). Aun así, a

pesar del avance se presentan desafíos en su implantación, como por ejemplo en la estandarización y la sostenibilidad económica de los equipos, aunque se observan soluciones que usan *blockchain* para validar comunicados oficiales y la colaboración descentralizada entre revisores (ProBlock, 2021).

Paralelamente, las APIs (interfaces de programas informáticos) de *fact-checking* automatizan procesos clave posibilitando que periodistas e instituciones accedan en tiempo real a verificaciones de hechos, tanto por texto como mediante búsquedas multimodales que incluyan imágenes y vídeos. El estándar ClaimReview, impulsado por Google y organizaciones aliadas, fortalece la integración de datos estructurados que los buscadores pueden analizar y contrastar, facilitando así el rastreo de afirmaciones verificadas en la web. Además, el auge de plataformas abiertas como Loki y WikiCheck, y la integración de bases científicas como Semantic Scholar, potencian el análisis automático y transparente de afirmaciones, garantizando no sólo velocidad sino también mejor desarrollo en los procesos de verificación (Factiverse, 2024). En España existe el proyecto Iveres, realizado entre RTVE y la U. Autónoma de Barcelona para VerificaRTVE, creado entre 2022 a 2024 e implantación durante 2025. Este sistema verifica frente a bases de datos de noticias ya existentes de medios fiables, así como de artículos académicos contrastados, la probabilidad de que una noticia aparecida en redes sociales sea verdadera o falsa (RTVE, 2023). Igualmente están en desarrollo otros modelos para empresas como Newtral o Maldita.

Entre las tendencias que se están llevando en la actualidad para mejorar la verificación de hechos destacan la automatización multimodal (capaz de analizar de forma unificada textos, sonidos e imágenes al mismo tiempo) y la estrategia de *prebunking* (estrategia preventiva que busca inmunizar a la ciudadanía ante la desinformación, alertándola de antemano sobre mensajes falsos o manipulativos antes de que estos circulen y se afiancen en el imaginario social). El empleo de inteligencia artificial en la generación y detección de desinformación, especialmente ante el auge de los *deepfakes*, exige un gran esfuerzo entre la capacidad tecnológica y el juicio humano, pues la precisión y confiabilidad de estos sistemas automatizados todavía presenta críticas y retos importantes. El futuro del *fact-checking* se perfila, por tanto, como un entramado híbrido donde la tecnología amplifica, pero no reemplaza, el papel de los profesionales de la información, cuya responsabilidad ética y crítica sigue siendo imprescindible.

2.7.3. Regulación, políticas públicas y el papel de las plataformas digitales

La Unión Europea ha construido el régimen regulatorio más ambicioso a nivel global para combatir la desinformación y los discursos de odio en el entorno digital. Este sistema se articula en torno a tres instrumentos principales que han evolucionado desde la autorregulación voluntaria hacia obligaciones jurídicamente vinculantes con sanciones económicas severas. El Reglamento (UE) 2022/2065 de Servicios Digitales (DSA) como piedra angular, establece un sistema escalonado de responsabilidades proporcionales al tamaño y riesgo de las plataformas. Con aplicación plena desde febrero de 2024, distingue entre servicios intermediarios, plataformas de alojamiento y plataformas de muy gran tamaño (VLOPs, con más de 45 millones de usuarios mensuales en la UE), que enfrentan obligaciones como auditorías anuales independientes, evaluaciones de riesgos sistémicos y medidas de mitigación proactiva.

Paralelamente al DSA, la UE ha desarrollado códigos de conducta específicos que operan como instrumentos de corregulación. El Código de Buenas Prácticas en materia de Desinformación, establecido en 2018 y reforzado en junio de 2022 con 34 signatarios, constituye el primer instrumento autorregulador de esta naturaleza a escala mundial. Establece 44 compromisos y 128 medidas específicas que incluyen transparencia en publicidad política, desmonetización de desinformación, herramientas de verificación de hechos y acceso a datos para investigadores, y aunque su adhesión sigue siendo voluntaria, el cumplimiento se considera medida adecuada de mitigación de riesgos y forma parte de la auditoría anual independiente a la que están sujetas las VLOPs. En enero de 2025 se lanzó el Código de Conducta para la lucha contra la incitación ilegal al odio en línea, que renueva los compromisos de 2016, y ha sido firmado por Meta, TikTok, YouTube, LinkedIn entre otras plataformas, reforzando la necesidad de actuación en 24 horas y establece mecanismos de monitoreo mediante investigadores acreditados.

España designó a la Comisión Nacional de los Mercados y la Competencia (CNMC) como Coordinadora de Servicios Digitales, responsable de supervisar el Cumplimiento. La implementación nacional ha enfrentado retrasos significativos, ya que aunque el Reglamento europeo es de aplicación directa desde febrero de 2024, España no ha completado a finales de 2025 las reformas legislativas necesarias para dotar a la CNMC de competencias y recursos específicos.

La legislación europea rompe, con estas normativas, con el principio de neutralidad absoluta de las plataformas (no son meros transmisores, sino que

también son responsables), estableciendo análisis de cómo sus sistemas y algoritmos se utilizan para difundir o amplificar contenidos incorrectos o engañosos. Una de las tensiones principales con estas empresas radica en el acceso a datos e información sobre funcionamiento algorítmico, donde la Comisión Europea ha iniciado investigaciones formales contra varias plataformas al establecer procedimientos de pago y diseños que disuaden del ejercicio de derechos, convirtiendo la transparencia en "promesa vacía" según denuncian muchos investigadores académicos.

Los informes de seguimiento evidencian deficiencias en el cumplimiento, ya que la *European Fact-Checking Standards Network* (EFCSN), que agrupa más de 50 organizaciones de verificación, ha alertado que más de un año y medio después de la firma del Código de 2022, la mayoría de las grandes plataformas están lejos de aplicar las medidas comprometidas, mientras que la propia Comisión Europea ha reconocido que siguen persistiendo la propaganda y desinformación automatizada a gran escala aunque sea denunciada (EDMO, 2023). En noviembre de 2025 la UE anunció la creación del Escudo Europeo de la Democracia para salvaguardar el espacio informativo de la desinformación y manipulación proveniente de fuera de sus estados miembros, mediante un sistema de alerta rápida gestionado por el Servicio Europeo de Acción Exterior (Comisión Europea, 2025).

Tal como se expone en el párrafo anterior, la desinformación y el odio persiste en las plataformas de redes sociales en internet aunque sean denunciadas en muchas ocasiones. Valga como muestra el experimento del Stratcom de la OTAN entre agosto y septiembre de 2024, en el que con un gasto total de 58€ se realizaron 44 publicaciones desinformativas, con 1.150 comentarios que los apoyaban, 11.725 me gustas, 3.150 compartidos y 8.233 vistas en Facebook, Instagram, Youtube, Tiktok, VKontakte y X. Dicho gasto fue 3 veces inferior a lo que había costado en 2023. A las 4 semanas solo se eliminó de media el 15% de las cuentas falsas identificadas, la tasa más baja de toda la serie (25% en 2021, 16% en 2022/23), y con unos resultados por plataforma: X eliminó el 50%, Facebook alrededor del 25%, TikTok un 3% y YouTube, Instagram y VK solo alcanzaron un 2%. En cuanto a la actividad no auténtica (me gusta, comentarios, visualizaciones, etc.), más del 93% de ese *engagement* falso comprado seguía activo. Tras la denuncia de esas cuentas falsas, en los cinco primeros días la tasa de eliminación sólo llegó hasta el 6% según la plataforma. Facebook fue la que más eliminó (6%), seguida de VK (≈3,3%), X, Instagram y YouTube se movieron entre el 1,3% y el 2,7%, y TikTok quedó en torno al 0,7% (Bergmanis-Korāts, & Haiduchyk, 2024).

2.8 RESUMEN DE LA SEGUNDA PARTE

A continuación se expone un esquema de lo dispuesto en esta segunda parte:

1. Planificación de la Campaña

- Definición de objetivos SMART: Se establecen metas específicas, medibles, alcanzables, realistas y limitadas en el tiempo. Los objetivos típicos incluyen degradar al adversario, desacreditar fuentes creíbles, distraer la atención o dividir a la sociedad.
- Segmentación y perfilado: Se clasifica a la población en grupos de interés basándose en el rastro digital de sus teléfonos móviles y redes sociales.
- Análisis de la audiencia objetivo:
  - Geográfico y demográfico: Identificación de áreas vulnerables (ej. barrios con tensiones sociales) y ajuste del mensaje por edad o género.
  - Psicográfico (modelo OCEAN): Clasificación por rasgos de personalidad (Apertura, Responsabilidad, Extraversión, Amabilidad y Neuroticismo) para crear mensajes hiper-personalizados.
  - Teoría de redes: Identificación matemática de nodos influyentes y cuentas que actúan como puentes para maximizar la difusión.

2. Preparación de la Campaña

- Desarrollo de narrativas (*storytelling*): Uso de la Pirámide de Freytag para estructurar la historia en cinco actos: Exposición, Acción ascendente, Clímax, Acción descendente y Desenlace.
- Tipología de narrativas: Aprovechamiento de vulnerabilidades, creación de narrativas conspirativas, de estigmatización o de deshumanización.
- Creación de contenidos y artefactos: Técnicas de manipulación con uso de medias verdades, cheapfakes, deepfakes, jajaganda, etc.
- Establecimiento de legitimidad: Creación de expertos falsos, webs de noticias ficticias o suplantación de identidades de confianza.

3. Distribución y Ejecución

- Técnica de astroturfing: Simulación de un movimiento espontáneo "desde abajo" (*grassroots*) mediante fases coordinadas:
  - Distribución: Cuentas nano-influencers inician el mensaje.
  - Amplificación: Captación de periodistas o grandes influencers para ganar credibilidad.

    - Inundación: Uso masivo de bots para saturar la conversación en horas pico.
- Cebado de bomba (*Pump Priming*) o globo sonda: Realización de ensayos controlados en pequeñas localidades (3.000-5.000 habitantes) para medir el impacto antes del lanzamiento masivo.
- Narrativas de contrarréplica: Uso de tácticas como el *whataboutism* (contraacusación), la defensa Chewbacca (argumentos sin sentido para confundir) o el Gish Gallop (aluvión de argumentos falsos) para neutralizar críticas.

4. Acción Offline y Monetización

- Salto al mundo real: Traslado del conflicto digital a manifestaciones o disturbios físicos mediante "especialistas de conversión" que transforman la emoción digital en violencia callejera.
- Modelo de negocio y monetización: Ingresos por publicidad programática, venta de merchandising, donaciones en criptomonedas o *crowdfunding*.
- Lavado de información: Proceso de tres etapas (emplazamiento, superposición e integración pública) para ocultar el origen de la mentira y que acabe en medios legítimos.

5. Evaluación e Infraestructura

- Medición de eficacia (KPIs): Análisis del alcance, el universo emocional generado y los cambios de comportamiento en los usuarios.
- Infraestructura tecnológica: Uso de granjas de trolls de cuarta generación (móviles virtuales en la nube) y compra de cuentas verificadas en el mercado negro.

6. Prevención y Defensa

- Alfabetización mediática: Desarrollo de pensamiento crítico que cuestione no solo el bulo, sino la arquitectura económica de la desinformación.
- Herramientas técnicas: Uso de blockchain para trazabilidad de noticias, APIs de fact-checking y estrategias de prebunking (inmunización previa).

3. Conclusiones

A lo largo de estas páginas ha quedado claro que la desinformación y el discurso de odio no son fenómenos recientes. Acompañan a la humanidad desde sus orígenes, aunque su forma, alcance y capacidad de influencia hayan cambiado profundamente. En poco más de un siglo, la manipulación informativa ha alcanzado cotas de sofisticación inéditas, combinando con maestría elementos sociológicos, psicológicos, comunicativos y tecnológicos. Sin embargo, igual que la desinformación o el odio no surgen de la nada, tampoco operan en el vacío, pues se alimentan de las fisuras ya abiertas en las sociedades, las amplifican y las convierten en escenarios de confrontación. Allí donde existen tensiones o conflictos latentes, estas fuerzas hallan terreno fértil. Los enfrentamientos artificiales difícilmente prosperan, pero los que explotan grietas reales, en cambio, resultan devastadores.

Durante las primeras décadas del siglo XX, los grandes arquitectos de la propaganda perfeccionaron sus métodos y descubrieron algo esencial: el poder de las emociones como motor de la conducta humana. Aunque tendamos a vernos como seres racionales, la experiencia cotidiana demuestra lo contrario, y más aún cuando actuamos en grupo. Las emociones guían nuestras respuestas, condicionan nuestros juicios y moldean nuestras decisiones. Por eso, quien logre dirigirse a ellas tiene muchas más probabilidades de influir con éxito que quien apela solo a la razón. La psicología, la comunicación y el marketing (tanto comercial como político) llevan un siglo confirmándolo y aprovechándolo con eficacia. Las primeras teorías comunicativas del siglo pasado pecaban quizá de simplistas, al suponer que todos reaccionaban igual ante los mensajes, pero sus resultados fueron asombrosos. Basta recordar la potencia propagandística de la Alemania nazi o las campañas de Edward Bernays, cuya influencia aún perdura en gestos tan cotidianos como el de desayunar tocino en el "típico" desayuno estadounidense.

Hoy la diferencia fundamental radica en la capacidad que tenemos (o más bien, que tienen sobre nosotros) para ser clasificados con una precisión inima-

ginable hace apenas unas décadas. A través de la huella que dejamos en internet, del uso cotidiano del teléfono móvil, de nuestras interacciones en redes sociales, de los videojuegos que jugamos, de la música que escuchamos o del equipo de fútbol al que seguimos, cada persona se convierte en un conjunto de datos perfectamente trazable. El mercado de la información personal se ha convertido ya en uno de los más lucrativos del planeta. Gracias a esta acumulación masiva de datos, los algoritmos son capaces de delinear con una exactitud inquietante el perfil psicológico de cada individuo. Esa capacidad equivale, en la práctica, a abrir la mente humana de par en par, pues permite identificar lo que tememos, lo que deseamos y lo que creemos, para manipular esos resortes profundos y empujarnos a comprar, votar o actuar de determinadas maneras.

Aunque estas técnicas aún no logran afectar a toda la población, ya resultan lo bastante eficaces como para modificar preferencias, comportamientos o resultados electorales con un impacto sin precedentes. La llegada de la inteligencia artificial y de los grandes modelos de aprendizaje automático (capaces de detectar patrones entre millones de datos) ha acelerado este proceso hasta hacerlo casi evolutivo. Durante la pandemia del Covid-19, por ejemplo, cerca del 60 % de la población mundial estuvo expuesta a desinformación de tipo conspirativo sobre las vacunas (del Campo Huerta, 2025).

La reacción más habitual ante este panorama suele ser la negación. Pensamos que la manipulación alcanza a los demás, no a nosotros. Creemos que quienes caen en esos engaños carecen de preparación o de criterio. Que nosotros, por ser personas racionales, no somos vulnerables a mensajes emocionales. Craso error. Las emociones actúan desde el núcleo más profundo de la mente y, cuando se logra activarlas, abren una puerta directa a la persuasión. Nadie queda completamente al margen de su influencia.

A lo largo de la historia, tanto las instituciones como las empresas o los estados han perfeccionado sus métodos de influencia sobre las masas. Ya no se trata solo de campañas diseñadas para un periodo electoral o de estrategias comunicativas improvisadas ante un hecho puntual. Aunque algunos estudios académicos cuestionan el alcance real de ciertas operaciones de desinformación (Martínez Ron, 2023), lo cierto es que muchas de ellas se despliegan con visión estratégica, a gran escala y a largo plazo. Bezmenov, antiguo agente del KGB, ya explicaba cómo los servicios soviéticos preparaban operaciones informativas cuyo efecto podía sentirse décadas después. Hoy ese modelo encuentra condiciones ideales con una tecnología más avanzada, datos personales en abundancia, y costes cada vez menores para diseñar y difundir mensajes de manipulación masiva.

El resultado es un ecosistema comunicativo propicio para la expansión de la desinformación, la polarización y el discurso de odio. Basta con observar la precariedad extrema de los trabajadores encargados de filtrar o verificar contenidos (muchos de ellos en países como Kenia, con sueldos de apenas tres euros por hora) para comprender el lado más invisible y cruel de esta industria (Bajo Erro, 2023). Las campañas que nacen en un territorio quizá no resulten efectivas en otro, por falta de conocimiento del contexto cultural y social, pero su evolución es cuestión de tiempo y de recursos. Con el desarrollo técnico y la globalización de los datos, cada vez menos fronteras impiden que la manipulación informativa encuentre nuevos espacios donde prosperar.

Un ejemplo ilustrativo de la evolución tecnológica aplicada a la persuasión política lo encontramos en las campañas electorales de Estados Unidos. En 2008, el equipo de Barack Obama marcó un punto de inflexión al aprovechar por primera vez el potencial de las redes sociales para conectar con distintos segmentos del electorado. Cuatro años más tarde, en 2012, su equipo dio un paso adelante incorporando análisis de sentimientos y estrategias de segmentación más afinadas, capaces de adaptar los discursos y narrativas a públicos cada vez mejor definidos.

En 2016, la campaña de Donald Trump llevó esa lógica a un nuevo nivel con el llamado Proyecto Álamo, que integró el trabajo de Cambridge Analytica y la colaboración directa (y en ocasiones opaca) de grandes empresas de redes sociales y análisis de datos. Aquella campaña consolidó la microsegmentación política, un método que permite diseñar mensajes específicos para grupos minúsculos de población, calibrados según su perfil emocional, ideológico o conductual ya no solo para convencer a favor, sino también para desincentivar el voto al contrario. Para 2020, las capacidades tecnológicas de ambos partidos ya se encontraban equilibradas, pero en 2024 se alcanzó una nueva frontera: la simulación de sociedades completas mediante inteligencia artificial. Los equipos de campaña comenzaron a probar sus mensajes dentro de modelos predictivos que anticipaban las posibles reacciones de los votantes antes incluso de ponerlos en circulación. Ninguno de estos avances inventó emociones o ideas nuevas, simplemente aprendieron a detectarlas, amplificarlas y convertirlas en herramienta política.

Combatir este fenómeno es cada vez más difícil. Por un lado, intervienen intereses económicos y geopolíticos de enorme peso, por otro, la sofisticación de las técnicas y el volumen de información disponible permiten conocer los procesos mentales de los grupos sociales con una precisión inquietante. De ahí la necesidad urgente de promover una alfabetización mediática efectiva que

forme a toda la población, no solo a los más jóvenes, y de fortalecer las instituciones dedicadas a la verificación de hechos. Las plataformas y redes sociales deben asumir la corresponsabilidad que les corresponde sobre el funcionamiento de sus algoritmos y los contenidos que difunden, tal como exige la legislación europea. Sin embargo, con frecuencia las medidas llegan tarde, o ni siquiera se aplican.

El problema, además, no puede resolverse sólo mediante respuestas reactivas o parciales. La prevención debe comenzar desde las etapas más tempranas del sistema educativo, como ya ocurre en varios países del norte de Europa, pero también debe abarcar la vigilancia de los flujos financieros que sustentan la desinformación, el control de los *proxies* (individuos o grupos que actúan, consciente o inconscientemente, al servicio de intereses desinformativos), y la fiscalización de las empresas que operan en este lucrativo mercado.

No se trata de limitar la libertad de expresión ni el derecho a la información, sino de proteger la democracia frente a quienes se valen de ella para socavarla. Frente a la propaganda, apagar el fuego con fuego solo debilita los valores que se pretenden defender, donde cualquier acción debe sostenerse sobre la ética y el respeto a los principios democráticos. Debemos defender la libertad de expresión y de prensa propias de una sociedad democrática, sin embargo, también debemos cuestionarnos si esa misma libertad debe amparar a quienes, bajo la apariencia de informar, difunden sólo propaganda destinada a causar daño y socavar precisamente la libertad que afirman proteger. El peligro es real: si no se actúa con decisión y visión a largo plazo, puede llegar un momento en que el daño sea irreversible. Desmontar una mentira siempre exige mucho más esfuerzo que crearla, de un orden de magnitud, y esa asimetría, como recordaba el inicio de este libro, es quizá la mayor dificultad de nuestro tiempo.

Agradecimientos

Este libro ha sido realizado a muchas horas de búsqueda de información de informes, trabajos académicos e informaciones periodísticas, así como el trabajo de datos y demás análisis realizados en los proyectos e investigaciones académicas. Pero también gracias a la ayuda de una manera u otra de mucha gente a la cual quiero agradecer el aportar ideas, conocimiento, confianza o empuje en llevar a cabo este proyecto.

Así, quiero dar las gracias a Virginia y H., que fueron de los primeros en decirme que esto se merecía un libro completo. Pero también a D., Leticia, Facundo, Ramón, Ovidio, Alberto, Ismael, Óscar, Carlos, Nel, Xurde, Pablo, Elías, Ignacio... que bien me aportaron ideas, datos, lecturas o repaso de lo escrito. Y obviamente a mi familia por aguantarme y apoyarme.

Glosario

Para una mejor comprensión de algunos términos y técnicas, se recomienda emplear el descrito en el Foro de Lucha contra la Desinformación de 2024, publicado por el Departamento de Seguridad Nacional de España, en su capítulo 1: https://acortar.link/5b1EvA

Bibliografía

Abdel Aziz, M. (2023). دخان الحرب مستمر.. معارك نفسية بأسلحة خفية في صراع إيران وإسرائيل (تحقيق [El humo de la guerra persiste: Batallas psicológicas libradas con armas ocultas en el conflicto Irán-Israel (Investigación)]. El Watan News. https://acortar.link/rq82HH

Abi-Habib, M. (2025, 24 noviembre). Rusian disinformation comes to Mexico, seeking to rupture U.S. ties. The New York Times. https://acortar.link/iRnfzz

Abramson, J. (2017, 24 enero). 'Alternative facts' are just lies, whatever Kellyanne Conway claims. The Guardian. https://bit.ly/3KlaXfw

A.G.P. (2023, 20 febrero). Eliminalia: la empresa española que gana millones blanqueando corruptos. El Español. https://bit.ly/493ZWcL

Alandete, D.E. (2022). Russian interference in the Catalan independence risis (2014-2022). EPP group in the european parlament. https://acortar.link/wQotPi

Algaba, A. (2025, 14 marzo). Quién es Liberum: La asociación «contra el estado abusador» que recurre todas las ZBE, incluida la de Donostia. Diario Vasco. https://acortar.link/arg1f5

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211-236. https://www.doi.org/10.1257/jep.31.2.211

Alliance4Europe. (n.d.). Alliance4Europe. https://alliance4europe.eu/

Aouragh, M. (2016). Hasbara 2.0: Israel's Public Diplomacy in the Digital Age. Middle East Critique. 25 (3), pp. 271-297. https://doi.org/10.1080/19436149.2016.1179432

Applebaum, A. (2020). El ocaso de la democracia: La seducción del autoritarismo. Editorial Debate.

Arce-García, S. (2022). El astroturfing como medio de influencia: el caso de las niñas y el ajedrez en Asturias. En Escandón-Montenegro, P., & Tejedor, S. (eds) Escenarios Digitales en la Comunicación, 263-280. Gedisa.

Arce-García, S., & Díaz-Campo, J. (2024). HAARP conspiracy: Analysis of its role in the 2023 Turkey & Syria earthquakes on Twitter. Estudios sobre el Mensaje Periodístico, 30(2), 323-333. https://doi.org/10.5209/esmp.95257

Arce-García, S., Martín-Jiménez, V., & Rodríguez-Fernández, L. (2025). Unveiling Hate Speech Dynamics: An Examination of Discourse Targeting the Spanish Meteorological Agency (AEMET). Social Inclusion, 13, 9291. https://doi.org/10.17645/si.9291

Arce-García, S., Rodríguez-Fernández, L., Establés Heras, M. J., García Marín, D., Marín García, B., Martín Jiménez, V., Pérez Curiel, C., Said-Hung, E., Salaverría Aliaga, R., & Wagner, A. (2024). 125 términos sobre desinformación. En Trabajos del Foro Contra las Campañas de Desinformación (pp. 9-39). Ministerio de Presidencia - Gobierno de España. https://doi.org/10.5281/zenodo.14540402

Arce-García, S., Said-Hung, E., & Mottareale, D. (2022). Astroturfing as a strategy for manipulating public opinion on Twitter during the pandemic in Spain. Profesional de la información, 31(3), e310310. https://doi.org/10.3145/epi.2022.may.10

Arce-García, S., Said-Hung, E., & Mottareale-Calvanese, D. (2023). Tipos de campaña Astroturfing de contenidos desinformativos y polarizados en tiempos de pandemia en España. Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes, 21(1). https://doi.org/10.7195/ri14.v21i1.1890

Arce-García, S., & Said-Hung, E. (2022). Astroturfing y debate político español desde las redes sociales, un estudio de caso. Sociología, Problemas e Praticas, 100, 107-124. https://doi.org/10.7458/SPP202210025549

Arce-García, S., Said-Hung, E., & Montero-Díaz, J. (2024). Unmasking coordinated hate: Analysing hate speech on Spanish digital news media. New Media & Society, 27(10), 5848-5868. https://doi.org/10.1177/14614448241259715

Arce García, S., Vila Márquez, F., & Fondevila i Gascón, J. (2021). Polarización en Twitter durante la crisis de la COVID-19: Caso Aislado y Periodista Digital. Revista De Comunicación, 20(2), 29-47. https://doi.org/10.26441/RC20.2-2021-A2

Arce-García, S., Vila, F. & Fondevila- Gascón, J.-F. (2022). Análisis del discurso de Twitter en los debates electorales de 2019 en España: un estudio algorítmico comparado. Communication & Society, 35(1), 45-61. https://doi.org/10.15581/003.35.1.45-61

Arévalo Salinas, A. I. (2014). El movimiento social 15-M de España y la promoción de la protesta a través de sus vídeos en YouTube. Historia y Comunicación Social, 19, 153–63. https://doi.org/10.5209/rev_HICS.2014.v19.45122

Arias, C. (2025, 16 julio). Desarticulada una red prorrusa que realizaba ciberataques a infraestructuras críticas en países europeos. Infobae. https://acortar.link/EMW1A5

Aro, J. (2020). Putin's trolls: On the front lines of the information war. Yale University Press.

Artuch, M. J. (2025, 22 mayo). La seguridad nacional atribuye a Rusia campañas de desinformación por la DANA para promover desconfianza ciudadana. RTVE. https://acortar.link/Pxah60

Assis, C. (2023, 15 marzo). Principales medios de Brasil tendieron a reproducir en sus titulares las falsedades de Bolsonaro sobre el COVID-19, según un estudio. Latam Journalism Review - Knight Center. University of Texas. https://acortar.link/dxdxmg

Avaaz (2020). How Facebook can Flatten the Curve of the Coronavirus Infodemic. https://bit.ly/4qVrqaP

Badillo, A., & Arteaga, F. (2024, febrero). El impacto de la desinformación en España. Iberifier. https://bit.ly/43rtkpI

Baja, J.P.P., & Borja II, R.E. (2025). Countering misinformation and disinformation in Philippine education: A systematic review of strategies, challenges, and the role of media and information literacy. International Journal of Research and Innovation in Social Science, 9(8), 7281-7288. https://acortar.link/hkgXX7

Bakir, V., & Briant, E.L. (2024). Techniques and transformation in the digital influence industry. En E. L. Briant & V. Bakir (Eds.), Routledge handbook of the influence industry (cap. 4). Routledge. https://doi.org/10.4324/9781003256878-5

Ballesteros, R.R. (2020, 1 enero). Alvise, el tuitero provocador de Ciudadanos que dejó la política sin indemnización. El Confidencial. https://tinyurl.com/bde8shwd

Balmaceda, A. (2025, 21 agosto). Advierten sobre la reactivación de los ataques del grupo pro-ruso NoName05716 en España. Cybersecurity News. https://acortar.link/yOVJhw

Bank, J., Stack, L., & Victor, D. (2018, 1 agosto). From 2018: Explaining QAnon, the Internet Conspiracy Theory That Showed Up at a Trump Rally. The New York Times. https://acortar.link/G4Umgb

Barbu, E., Banerjee, S., Lim, T., & Zīvere, L. (2025, 16 diciembre. AI in support of StratCom: The use and evaluation of large language models in less widely used official EU languages. NATO StratCom CoE. https://acortar.link/buM9BE

Barrouquere, B. (2023). El Paso shooting suspect may have authored manifesto containing white nationalist talking points. Southern Poverty Law Center. https://acortar.link/C5Z3Rg

BBC News (2017, 14 agosto). Charlottesville: Trump under fire for response to violence. https://www.bbc.com/news/world-us-canada-40915569

BBC News (2023, 19 mayo). Carole Cadwalladr ordered to pay £1.2m costs in Arron Banks libel trial. https://acortar.link/MOSDkw

Bencherif, A., & Carignan, M.-E. (2023). Exploratory research report on the information environment in a political and security crisis context in the Sahel Region. StratcomCoe OTAN. https://acortar.link/xKbKJs

Beran, D. (2019). It Came from Something Awful: How a Toxic Troll Army Accidentally Memed Donald Trump into Office. St. Martin's Press.

Bergmanis-Korats, G., Bertolin, G., Puzule, A., & Zeng, Y. (2024). AI in Support of StartcCom Capabilities. NATO Strategic Communications Centre of Excellence. https://bit.ly/3W4rUNW

Bergmanis-Korāts, G., & Haiduchyk, T. (2024). Social media manipulation for sale: Experiment on platform capabilities to detect and counter inauthentic social media engagement. NATO Strategic Communications Centre of Excellence. https://acortar.link/g3h7PG

Bernays, E.L. (1928). Propaganda. Horace Liveright.

Bernays, E.L. (1955). The engineering of consent. University of Oklahoma Press.

Bhuiyan, J., & Levin, S. (2023, 6 septiembre). Así utiliza EE.UU. perfiles falsos en redes sociales para investigar a inmigrantes. El Diario. https://acortar.link/2TaNqM

Biescas, A., Allen, D.G., & Hernández, S. (2024, 10 julio). Enjambre de desinformación favorable al Kremlin en español. EFE Verifica. https://acortar.link/cDWC2c

Blanco, P.R. (2017, 10 diciembre). Así arruinaron los 'trolls' rusos la vida de Jessikka Aro: La periodista finlandesa se convirtió en el objetivo de una campaña de difamación tras investigar los perfiles falsos de las redes sociales que difunden propaganda a favor del Kremlin. El País. https://acortar.link/18AX13

Blanco, P.R., & Abril G. (2023, 12 febrero). Noticias en suajili y series comunistas: así ha conquistado China el espacio mediático en África. ElPais.com. https://acortar.link/KBqshC

Blatny J.M., & Søndergaard, S. (2025, 19 diciembre). Cognitive Warfare. NATO Science and Technology Organization-STO. https://acortar.link/7KkCLb

Bloomberg Línea. (2021, 2 diciembre). ¿Por qué Twitter ha eliminado más de 270 cuentas afines a Nicolás Maduro? https://bit.ly/4hUePkc

Bolaños Noguera, M.A., Zambrano Hinestrosa, J.R., & Tumal Chaves, O. A. (2024). Incidencia de las redes sociales en el pensamiento crítico en estudiantes de educación superior. Revista Huellas, 10(2). https://doi.org/10.22267/huellas.24102.22

Bolt, N., & Lange-Ionatamishvili, E. (2026). The Nextgen information environment. *Nato Stratcom Coe*. https://acortar.link/vy8KiU

Borràs Rius, A. (2025, 28 enero). El mapa que explica la influencia rusa en España. Agenda Pública. https://acortar.link/9XWQYw

Bot Ruso (2019). Confesiones de un Bot Ruso. Debate.

Bradshaw, S., Bailey, H., & Howard, P.N. (2020). Industrialized Disinformation 2020 Global Inventory of Organized Social Media Manipulation. Computational Propaganda Research Project. Oxford Internet Institute. https://acortar.link/W93cst

Bradshaw, S., & Howard, P.N. (2019). Inventario global de la manipulación organizada de redes sociales 2019. Universidad de Oxford. https://acortar.link/wMxRX7

Brandolini, A. [@ziobrando]. (2013, 11 enero). The bullshit asymmetry: the amount of energy needed to refute bullshit is an order of magnitude bigger than to produce it [Tweet]. Twitter. https://acortar.link/2Msv8V

Briant, E.L. (2020). Propaganda Machine. Inside Cambridge Analytica and the Digital Influence Industry. https://www.propagandamachine.tech/

Briant, E.L., & Jones, M.O. (2025). A century of propaganda studies: from pen and sword to surveillant smartphone. Critical Studies in Media Communication, 42(1), 64–68. https://doi.org/10.1080/15295036.2025.2464184

Buziashvili, E., & et al. (2024, junio). Another battlefield: Telegram as a digital front in Russia's war against Ukraine. DFRLab - Atlantic Council. https://acortar.link/N5Ev1J

Byung-Chul, H. (2018). Infocracia. La digitalización y la crisis de la democracia. Anagrama.

Cadwalladr, C., & Graham-Harrison, E. (2018, 17 marzo). Cambridge Analytica: links to Moscow oil firm and St Petersburg university. The Observer. https://acortar.link/vSnwgB

Caldwell, W.B., IV, Murphy, D.M., & Menning, A. (2009, 26 mayo). Learning to leverage new media: The Israeli forces in recent conflicts. Military Review US Army. https://acortar.link/5upOpg

Calvillo, D., León, A., & Rutchick, A.M. (2024). Personality and misinformation. Current opinion on psychology, 55, 101752. https://doi.org/10.1016/j.copsyc.2023.101752

Campbell, W. J. (2001). Yellow journalism: Puncturing the myths, defining the legacies. Praeger

Cañizález, A. (2021). Desinformación en Venezuela: Reflexiones en tiempos de pandemia. Universidad Católica Andrés Bello. https://acortar.link/5E5RJf

Carrasco Rodríguez, B. (2021). Information laundering in Germany. NATO Strategic Communications Centre of Excellence. https://bit.ly/48VMkAl

Carrasco Rodríguez, B. (2023). Information laundering in the Nordic-Baltic region. NATO Strategic Communications Centre of Excellence. https://bit.ly/49idilZ

Casquinho, M., Vasconcelos, A., Moreno, J., Cardoso, G., Palma, N., Paisana, M., Pinto-Martinho, A. (2024). Europeias 2024 - Amplificação do discurso político online e desinformação em Portugal. Publicações OberCom http://bit.ly/4mkvYUT

Castells, M. (2009). Comunicación y poder. Alianza Editorial.

Cazadores de Fake News. (2021, 18 enero). Sin RT no hay paraíso: ¿cómo funciona la maquinaria de propaganda en Twitter de Nicolás Maduro? https://acortar.link/SDAEJF

Cellhasher (2025). The original Phone Farm Box. https://cellhasher.com/

CESIE (2022, 31 enero). What is the true cost of disinformation? https://acortar.link/B3PbXl

Châtelet, V., & Lesplingart, A. (2025, 18 abril). Russia's Pravda network in numbers: Introducing the Pravda Dashboard. DFRLab. https://acortar.link/tvOj9G

Colas, X. (2017, 4 octubre). La conexión moscovita del 'procés' con los hackers rusos. El Mundo. https://acortar.link/PavJCM

Colom Piella, G. (2020). Anatomía de la desinformación rusa. Historia y comunicación social, 25(2), 473-480. https://doi.org/10.5209/hics.63373

Colon, D. (2025). La guerre de l'information: les États à la conquête de nos esprits. Éditions Tallandier.

Comisión Europea (2025, 12 noviembre). El Escudo Europeo de la Democracia y la Estrategia de la UE para la Sociedad Civil allanan el camino hacia unas democracias más fuertes y resilientes. https://acortar.link/H4Xuxv

Coface (2022). Coface utiliza blockchain para certificar sus comunicados de prensa. https://acortar.link/pBmqyn

Correo del Caroní (2022, 7 marzo). Probox reporta protesta digital de tuiteros de la patria porque el gobierno no les paga. https://acortar.link/Imh610

Cubero Trujillo, I. (2020). 4TP: Hacia una Cuarta Teoría Política. Alexander Dugin y el Neoeurasianismo. Tiempo Devorado, 6(1), 3–15. https://doi.org/10.5565/rev/tdevorado.145

Cuesta, J.G. (2025, 10 noviembre). El Kremlin toma el control absoluto sobre el internet de Rusia. ElPais.com. https://bit.ly/47TKJsC

da Empoli, G. (2019). Los ingenieros del caos. Vestígio.

Dawson, M., & Innes, P. (2021). Political Quarterly. Cardiff University. https://acortar.link/OTX6zK

del Campo Huerta, A.M. (2025, 9 septiembre). Política y desinformación: la salud global en riesgo. Observatorio de Medios Digitales del Tecnológico de Monterrey. https://bit.ly/41LwS5q

del Castillo, C. (2022, 7 abril). Hackers vinculados a Bielorrusia intentan controlar perfiles de militares ucranianos para publicar bulos. eldiario.es. https://bit.ly/49MQZ7X

del Castillo, C. (2023, 6 de julio). Rusia ensaya la desconexión total de Internet global. El Diario. https://acortar.link/QgZD0u

Democracy Now (2016, 14 octubre). Presentador de Fox Lou Dobbs se disculpa tras publicar información personal de Jessica Leeds. https://bit.ly/4nMXEmj

Departamento de Seguridad Nacional-DSN (2024). Informe Anual de Seguridad Nacional 2024. https://bit.ly/4oMcpXd

Deutsche Welle. (2023, 7 mayo). Russia staging protests for anti-Ukraine propaganda: report. DW. https://bit.ly/4nvC2um

de Pedro, N., & et alters. (2023). Estudio de la desinformación rusa a nivel internacional. En Foro de lucha contra la desinformación (Ed.), Lucha contra las campañas de desinformación en el ámbito de la seguridad nacional: propuestas de la sociedad civil (pp. 48-65). Departamento de Seguridad Nacional, Presidencia del Gobierno de España. https://acortar.link/xfGdrl

Dippel, J., Dupré la Tour, M., Niu, A., Roy, S., & Vetta, A. (2024). Eliminating majority illusion is easy. https://arxiv.org/pdf/2407.20187

Disarm Foundation. (n.d.). Disarm Foundation. https://www.disarm.foundation/

Disarm Framework. (n.d.). Disarm Framework. https://disarmframework.herokuapp.com/

Disinfo Africa. (2023). New cold war: Russia's self-defence narrative is the winner in Egypt. https://acortar.link/3zkxQW

Disinfo Africa (2025, 3 julio). AI exposes fake accounts manipulating Wikipedia content. https://acortar.link/6FGZ9S

DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., Matney, R., Fox, R., Albright, J., & Johnson, B. (2019). The Tactics & Tropes of the Internet Research Agency. Congress of the United States. https://bit.ly/3PNCRz5

DSN - Presidencia del Gobierno (2025). Informe anual de Seguridad Nacional. https://acortar.link/1Iwq8o

Dsouza, S., & Jones, M. O. (2024). The Qatar plot report. https://sites.google.com/view/theqatarplotreport/

Donovan, J., Dreyfuss, E., & Friedberg, B. (2022). Meme Wars: The Untold Story of the Online Battles Upending Democracy in America. Bloomsbury Publishing.

Dumont, E., Solis, J., & Zaleski, L.,(2023). North Macedonia: Profile of Media Ownership and Potential Foreign Influence Channels. Williamsburg, VA: AidData at William & Mary.https://acortar.link/sUxJI6

Duncan, P., Hernandez, R., Morresi, E., Gutiérrez, P., Blight, G., & McMullan, L. (2025, 28 septiembre). Inside the everyday Facebook networks where far-right ideas grow. The Guardian. https://bit.ly/42J1qFv

de Gregorio Vicente, O. (2023). Desarrollo del Pensamiento Crítico: Nuevas tecnologías y Redes Sociales (Trabajo fin de estudio). Universidad Internacional de La Rioja.

de Pedro, N., & et alters. (2023). Estudio de la desinformación rusa a nivel internacional. En Foro de lucha contra la desinformación (Ed.), Lucha contra las campañas de desinformación en el ámbito de la seguridad nacional: propuestas de la sociedad civil (pp. 48-65). Departamento de Seguridad Nacional, Presidencia del Gobierno de España. https://acortar.link/xfGdrl

Di Domenico, G., & Ding, Y. (2023). Between brand attacks and broader narratives: How direct and indirect misinformation erode consumer trust. Current Opinion in Psychology, 54, 101716. https://doi.org/10.1016/j.copsyc.2023.101716

EDMO. (2023). No violations found: Europe's digital safety law fails when users report content. https://acortar.link/CaCGny

Elkins, K. (2025). Beyond plot: How sentiment analysis reshapes our understanding of narrative structure. Journal of Cultural Analytics, 10(3). https://doi.org/10.22148/001c.143671

Elley, B. (2021) The rebirth of the West begins with you!"—Self-improvement as radicalisation on 4chan. Humanities and Social Sciencies Communication, 8, 67. https://doi.org/10.1057/s41599-021-00732-x

El Economista (2024, 16 abril). Campañas antivacunas recaudaron más de 100 millones de dólares por difundir información falsa. El Economista. https://bit.ly/47H2H1u

El equipo de campaña (n.d.). Elígenos. https://elequipo.com/eligenos/

El País (2018, 21 marzo). La huella mexicana de Cambridge Analytica. El Pais.com. https://acortar.link/2eCGNv

El País (2024, 28 agosto). La justicia francesa imputa al fundador de Telegram y lo deja en libertad bajo fianza. https://acortar.link/ohKXag

Emergui, S. (2023, 15 febrero). El 'Equipo Jorge': interferencia en campañas electorales y ciberataque en la consulta independentista catalana del 2014. El Mundo. https://acortar.link/hQR7ig

Ennis, G. (2023). Dark PR: How Corporate Disinformation Harms Our Health and the Environment. Daraja Press.

Epifanova, E. (2020, 16 enero). Deciphering Russia's "Sovereign Internet Law". German Council on Foreign Relations (DGAP). https://acortar.link/CzOMUg
Equipo de Investigación (2021, 15 enero). Equipo de Investigación logra contactar con la médica negacionista Natalia Prego: "Hagan examen de conciencia". La Sexta. https://bit.ly/493wIe8
Escribano, M. (2023, 28 agosto). La vigilancia en Internet: el gobierno de EE.UU. y la inteligencia artificial. El Confidencial. https://acortar.link/YFCC1F
Espaliú-Berdud, C. (2023). Use of disinformation as a weapon in contemporary international relations: accountability for Russian actions against states and international organizations. Profesional de la información, 32(4), e320402. https://doi.org/10.3145/epi.2023.jul.02
Euronews (2025, 17 julio). Desarticulada la red de hackers prorrusos NoName 057, que perpetró ciberataques en España. Euronews. https://acortar.link/KMGuvf
Euronews (2025a, 22 mayo). Seguridad Nacional confirma que Rusia explotó la tragedia de la DANA. Euronews. https://bit.ly/3JxTMaj
Euronews (2025b, 9 enero). Euroverify: Las fotos de una morgue no muestran los cuerpos ocultos de las víctimas de la DANA. Euronews. https://bit.ly/3LgUk5e
European Court of Auditors (2021). Disinformation: The EU's response to information manipulation. https://acortar.link/YfJM4Z
European Union External Action (2024, junio). Doppelganger strikes back: FIMI activities in the context of the EE24. https://acortar.link/vNV5Bt
European External Action Service (2022). EEAS annual report. https://acortar.link/safvWv
European External Action Service (2023). EEAS Stratcom annual report 2023. https://acortar.link/BBnNuk
European External Action Service (2024). EEAS 2nd Report on FIMI Threats: Towards a framework for networked defence. European External Action Service. https://acortar.link/b2lBp4
European External Action Service (2024, abril). Operation False Façade: Insights from a FIMI information laundering scheme. EUvsDisinfo. https://acortar.link/nNEO7h
European External Action Service (EEAS) (2025). 3rd EEAS Report on Foreign Information Manipulation and Interference Threats. https://acortar.link/WmkjEg
Europol (2025). Desmantelamiento de un servicio de ciberdelincuencia: 7 detenidos. https://acortar.link/FRz20R
EUvsDisinfo (2024, 21 mayo). Grandes modelos de lenguaje: un nuevo frente en la guerra informativa rusa. EUvsDisinfo. https://bit.ly/480zKP7
EUvsDisinfo. (2025, 28 abril). Echoes of influence: Inside Russia's FIMI activities in Africa. Servicio Europeo de Acción Exterior (SEAE). https://acortar.link/bK7zmg
EUvsDisinfo (2025, 22 abril). A glossary: Who is who in the FIMI zoo? https://bit.ly/4lRsx81

Every-Palmer, S., Cunningham, R., Jenkins, M., & Bell, E. (2020). The Christchurch mosque shooting, the media, and subsequent gun control reform in New Zealand: a descriptive analysis. Psychiatry, Psychology and Law, 28(2), 274–285. https://doi.org/10.1080/13218719.2020.1770635

Factiverse (2024). Enhancing Fact-Checking with Semantic Scholar API. https://acortar.link/IPlyPt

Fernández, J.J. (2024, 22 julio). El Ejército tiene identificado un "ecosistema de desinformación" ruso en España con 179 altavoces. El Periódico. https://acortar.link/rGCB1L

Fernández, J.J., & Calleja Flórez, T. (2025, 9 junio). Pérez Dolset, implicado con Leire Díez en el complot contra la UCO, facilitó a PP y PSOE un programa de influencia electoral diseñado en Rusia. El Periódico. https://bit.ly/3IdLq75

Fernández Chapou, M. (2025, 3 septiembre). Infiltración de desinformación pro-Kremlin en la inteligencia artificial, un desafio global. Observatorio de Medio Digitales del Tecnológico de Monterrey. https://acortar.link/9NYA2o

Fernández Huerta, J. (1959). Medida sencillas de lecturabilidad. Consigna, 214, 29-32.

Flashpoint Intel Team (2023, 19 abril). Understanding Russia's "Sovereign Internet": What Happens If Russia Isolates Itself from the Global Internet? Flashpoint. https://acortar.link/rEMFhD

France24. (2025, diciembre 24). Five Europeans denied US visas for combating hate speech online, accused of censoring 'American viewpoints'. France 24. https://acortar.link/nRQXLM

Frias Deniz, A. (2025, 3 octubre). ¿México, colonizado por propaganda rusa?. Observatorio de Medios Digitales del Tecnológico de Monterrey. https://acortar.link/Jyao8u

Fuente Cobo, I. (2025, 4 junio). La obsesión rusa: desinformación y propaganda en el Sahel. Ministerio de Defensa - CESEDEN. https://acortar.link/qGzm5X

Fulde-Hardy, M. (2023). Falsos amigos: La desinformación en la era de las redes sociales. Graphika. https://acortar.link/6QxbU5

Gagandeep (2025, 2 julio). Playing with hate: How far-right extremists use Minecraft to gamify radicalisation. The Global Network on Extremism and Technology (GNET). https://acortar.link/SYnrZb

Galán Cordero, C., Arroyo Guardeo, D., de Pedro, N., Gómez García, P., González Nagore, P., Manuel Pérez Triana, J., Marchal González, N., Mier y Teran, J., Pérez Bes, F., Portillo, I., Tun Navarro, J. F., & Valencia Martínez de Antoana, J. (2024). Capítulo 3: Trabajos Foro. Campañas de desinformación. Iniciativas 2024 (pp. 4–64). Departamento de Seguridad Nacional. Ministerio de Asuntos Exteriores, Unión Europea y Cooperación, Comisaría General de Información, Cuerpo Nacional de Policía. https://bit.ly/48OdaKH

Galeotti, M. (2019). The mythical 'Gerasimov Doctrine'and the language of threat. Critical Studies on Security, 7(2), 157-161. https://doi.org/10.1080/21624887.2018.1441623

Galeotti, M. (2022). Las guerras de Putin: De Chechenia a Ucrania. Desperta Ferro.

Galeotti, M. (2014, 21 abril). Putin's Empire of the Mind. How Russia's president morphed from realist to ideologue and what he'll do next. Foreign Policy. https://bit.ly/49NMgmu

García, J. (2021, 9 mayo). Metiendo la "gambita" con las damas: así es la discriminación de género en el ajedrez. El Confidencial. https://bit.ly/3L8X3O4

García-Estévez, N., Ballesteros-Aguayo, L., & Colussi, J. (2025). Desinformación y manipulación de la opinión pública: una revisión sistemática sobre astroturfing (2004-2024). Revista De Comunicación, 24(2), 159–181. https://doi.org/10.26441/RC24.2-2025-3988

García Márquez, J. M. (2010). El triunfo del golpe militar: el terror en la zona ocupada. En Francisco Espinosa Maestre, ed. Violencia roja y azul. España, 1936-1950. Crítica. pp. 81-150.

García-Marín, D. (2024). Periodismo contra la desinformación. Proceso y estructura de las verificaciones en el fact-checking. Infonomy, 2(2), e24026. https://doi.org/10.3145/infonomy.24.026

Gardner, D. (2025, 20 noviembre). Vance urged Bezos to make The Washington Post even more MAGA. The Daily Beast. https://acortar.link/IDxkWT

Gellman, B., & Poitras, L. (2013, 6 junio). U.S. intelligence mining data from nine U.S. Internet companies in broad secret program. The Washington Post. https://bit.ly/49Po8jp

Genfarmer (2025). Automation Mobile No-Coding for enterprise and Marketing Campaigns. https://genfarmer.com/

Gil, J., & Irujo, J.M. (2023, 15 febrero). Un empresario israelí se atribuye el ciberataque a la Generalitat en la consulta del 9-N de 2014. El País. https://acortar.link/N4J3w0

Gil, J., & Irujo, J.M. (2023, 17 febrero). Gobernadores mexicanos, narcos y jerarcas chavistas ficharon a una firma española de desinformación para lavar su imagen en la red. El País. https://bit.ly/4oT2a39

Gil, J., & Irujo, J.M. (2023, 17 febrero). Gobernadores mexicanos, narcos y jerarcas chavistas ficharon a una firma española de desinformación para lavar su imagen en la red. El País. https://bit.ly/4oT2a39

Gil, J., & Irujo, J.M. (2023b, 17 febrero). Eliminalia: Una firma española limpia en la red la imagen de corruptos, abusadores y narcos de 54 países. El País. https://bit.ly/4hGClB4

Gilbert, D. (2022, 4 abril). Inside Cyber Front Z, the 'People's Movement' Spreading Russian Propaganda. VICE. https://acortar.link/bSI9hT

Globalamericans. (2021). Measuring the impact of misinformation, disinformation, and propaganda in Latin America. https://acortar.link/8F1otN

Gold, M. (2017, 17 marzo). The Mercers: A dynasty of influence in American politics. The Washington Post. https://acortar.link/X6lufI

González, M. (2023, 3 noviembre). Un juez condena a Julio Ariza, patrón mediático de Vox, a pagar 4,5 millones por su gestión de Intereconomía Televisión. El País. https://tinyurl.com/2zy8pp9y

Gozálvez-Pérez, V., Valero-Moya, A., & González-Martín, M.-R. (2022). El pensamiento crítico en las redes sociales. Una propuesta teórica para la educación cívica en entornos digitales. Estudios Sobre Educación, 42, 35-54. https://doi.org/10.15581/004.42.002

Granovetter, M.S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360-1380.

Gragido, W., & Pirc, J. (2011). Seven Communalities of subversive multivector threats. En. Cybercrime and Espionage. An analysis of subversive multi-vector threats. (pp. 153-175). https://doi.org/10.1016/B978-1-59749-613-1.00009-1

Graphika Atlas (2025, marzo). Chinese state influence: Tariffs to tension. https://acortar.link/tUGIIM

Grazda, B. (2024). Weaponizing WhatsApp: A Case Study of a Digitally Amplified Institutionalized Riot System in India. En E. L. Briant & V. Bakir (Eds.), Routledge Handbook of the Influence Industry (Cap. 16). Routledge.

Greene, S., Soldatov, A., & Borogan, I. (2024). War Without End: Russia's Shadow Warfare. Center for European Policy Analysis (CEPA). https://acortar.link/fUMXM9

Grima, C. (2022). ¡Que las matemáticas te acompañen!. Ariel.

Grohmann, R., & Ong, J.C. (2024). Disinformation-for-hire as everyday digital labor: Introduction to the special issue. Social Media + Society, 10(1), 1–9. https://doi.org/10.1177/20563051231224723

Герасимов, В.В. (2022, 16 marzo). Ценность науки в предвидении [El valor de la ciencia en la previsión]. ВПК Новости. https://bit.ly/3XkKRML

González, M. (2025, 22 mayo). El informe de seguridad nacional atribuye a Rusia campañas de desinformación por la DANA. El País. https://acortar.link/g3w3Oc

Guerrero-Saade, J.A., Moore, D., Raiu, C., & Rid, T. (2017). Penguin's moonlit maze. Kaspersky Lab - Great - King's College London. https://ridt.co/d/jags-moore-raiu-rid.pdf

Gutierrez, N. (2018, 26 febrero). Bots, Assange, an alliance: Has Russian propaganda infiltrated the Philippines?. Rappler.com https://bit.ly/3RrD1Na

Gutiérrez, O. (2024, 13 mayo). Rusia culmina con éxito su operación de propaganda y desinformación en el Sahel. ElPais.com. https://acortar.link/tHDKv4

Hagey, K. (2021, 19 mayo). Peter Thiel, J.D. Vance invest in Rumble video platform popular on political right. The Wall Street Journal. https://acortar.link/fFnzwV

Hapal, D.K. (2024, 1 noviembre). The Philippines' disinformation machine. New Internationalist. https://acortar.link/8LzSAd

Hatemedia (2025). Observatorio digital de odio. https://hatemedia.es/

HatemediaReligion (2025). Análisis del odio religioso. https://hatemedia.es/objetivos-hatemedia-religion/
Hernando, A. (2020, 8 noviembre). Los bots llenaron Twitter de mentiras en las elecciones que ganó Donald Trump. Agencia SINC. https://acortar.link/rqKrOG
Hindman, M. (2018, 10 abril). Cómo funcionaba el modelo de Cambridge Analytica, según la persona que lo construyó. ElPais.com. https://acortar.link/7KrRlE
Holyoke, G. (2025, 25 abril). ¿Qué debe hacer Europa mientras Rusia gana influencia en el Sahel africano? Euronews.com. https://acortar.link/tVFl5e
Huang, A. (2025). Combatting and defeating Chinese propaganda and disinformation: A case study of Taiwan's 2020 elections. En M. Echeverría, S. G. Santamaría, & D. C. Hallin (Eds.), State-sponsored disinformation around the globe: How politicians deceive their citizens (pp. 121–135). Routledge. https://doi.org/10.4324/9781032632940-10
Hughes, A., & Cvetkovska, S. (2021). The Macedonian fake news industry and the 2016 US election. PS: Political Science & Politics, 54(1), 19-23.
Hutchinson, A. (2022, 17 mayo). Meta Shares Latest Numbers on Rules Enforcement, Including Abuse and Terror-Related Content and Fake Profiles. Socialmediatoday.com. https://bit.ly/4m7j4Jy
House of Commons (1925, 3 marzo). Subversive Propaganda (Great Britain and the Empire). Hansard Parliamentary Debates. https://bit.ly/4qQc0of
Infantes Capdevila, G. (2024, 31 marzo) De la 'jajaganda' al odio: la propaganda pro-Kremlin convierte en meme a un ucraniano mutilado. Newtral.es. https://bit.ly/3KlhSp2
Infobae (2020, 8 junio). Red AMLO: el "ejército de bots" que estaría detrás de la información pro López Obrador en las redes sociales. https://acortar.link/UZGqep
Insikt Group (2019, 30 septiembre). Disinformation service campaigns. Recorded Future. https://acortar.link/38WjT5
Institute for Strategic Dialogue (2025, 13 mayo). Investigation: How Russia Today is evading sanctions and spreading pro-Kremlin propaganda in Italy. https://acortar.link/AoEeDg
Institute for Strategic Dialogue, Alliance4Europe, Debunk.org, GMF Alliance for Securing Democracy, DEN Institute, EU DisinfoLab. (2023). Country Report: Assessment of Foreign Information Manipulation and Interference (FIMI) in the 2025 German Federal Election. https://acortar.link/ZZHHBl
Instituto CEU de Estudios Históricos (2012). Checas de Madrid. https://acortar.link/Expckd
Instituto Español de Estudios Estratégicos (2021). Desinformación y subversión (2.0) Las técnicas de la Guerra Fría reaparecen en el dominio informativo del siglo XXI. https://bit.ly/47KUzNd
Institute for the Study of War (2015, 21 septiembre). Putin's information warfare in Ukraine: Soviet origins of Russia's hybrid warfare. https://acortar.link/b1g4QG

Institute for the Study of War (2025, 30 junio). A primer on Russian cognitive warfare. https://acortar.link/0Fq9sd

Iriarte, D. (2024, 19 agosto). Habla el cronista de la guerra de la desinformación en México: "Somos el laboratorio del mundo". El Confidencial.com. https://bit.ly/3K1XgCh

Iriarte, D. (2025). Guerras cognitivas: Cómo estados, empresas, espías y terroristas usan tu mente como campo de batalla. Arpa Editores.

Jankowicz, N. (2020). How to Lose the Information War: Russia, Fake News, and the Future of Conflict. I.B. Tauris

Jeangène Vilmer, J.-B., Escorcia, A., Guillaume, M., & Herrera, J. (2018). Les manipulations de l'information: un défi pour nos démocraties (Rapport du Centre d'analyse, de prévision et de stratégie [CAPS] du ministère de l'Europe et des Affaires étrangères et de l'Institut de recherche stratégique de l'École militaire [IRSEM] du ministère des Armées). Paris. http://bit.ly/4pqCjRo

Jones, M.O. (2023). Nobody talks like this: Meet 50 AI. Dysinfluence - Substack. https://acortar.link/1slP2q

Jones, M.O. (2024, 24 julio). The Qatar plot: How a covert influence campaign helped Europe's far right. Al Jazeera. https://acortar.link/K1POnZ

Jones, M.O. (2025). Deception supply chains in the middle east, 2010–2023: A playbook of social media manipulation, disinformation and influence operations. En E. L. Briant & V. Bakir (Eds.), Routledge handbook of the influence industry (pp. 234–250). Routledge. https://doi.org/10.4324/9781003256878-13

Jones, M.O. (2025a). The $8 Million Influence Machine: Inside Israel's Recent U.S. Propaganda Campaigns. Dysinfluence. https://acortar.link/DsbQJP

Jones, M.O. (2025b). Search for Your Church (And see if it has been targeted by the Israeli Gov). Dysinfluence. https://acortar.link/U4yLTq

Jousset, A. (2022). Fuentes: Prigozhin y los mercenarios del Grupo Wagner [Documental]. CAPA PRESSE; ARTE. https://www.arte.tv/es/videos/113682-001-A/fuentes/

Kaiser, B. (2019). La dictadura de los datos. La verdadera historia desde dentro de Cambridge Analytica y de cómo el Big Data, Trump y Facebook rompieron la democracia y cómo puede volver a pasar. HarperCollins Ibérica, S.A.

Kalmoe, N.P. (2014). With passion and principle: Trade-offs and strategy in political argument. University of Chicago Press.

Kasianenko, K., & Boichak, O. (2024). Canonizing online activism: Memetic iconography in the North Atlantic Fella Organization. Media, War & Conflict, 18(2), 179-196. https://doi.org/10.1177/17506352241279957

Keller, F.B., Schoch, D., Stier, S., & Yang, J. (2019). Political Astroturfing on Twitter: How to Coordinate a Disinformation Campaign. Political Communication, 37(2), 256–280. https://doi.org/10.1080/10584609.2019.1661888

Kirchgaessner, S., Ganguly, M., Pegg, D., Cadwalladr, C., & Burke, J. (2023, 15 febrero). El mercado de desinformación de un equipo secreto israelí que vende servicios para

intervenir en elecciones en todo el mundo. Eldiario.es. https://acortar.link/oL3LRy
Lazarus, S. (2018). The Nigerian Cyber Fraudsters (Yahoo Boys) and Hip Hop Music in Nigeria: Mapping the Track. Criminology, criminal justice, 19(2), 63-80. https://acortar.link/QpuZ5N
La Moncloa (2025, 18 junio). ¿Qué son los discursos de odio y cómo podemos combatirlos?. http://bit.ly/3WJk64B
La Sexta (2020, 24 agosto). El negocio de la guerra: la empresa de mercenarios Blackwater podría expandirse por todo el mundo. https://acortar.link/wwEsgi
Le Monde (2023, 7 mayo). How Russia is staging fake protests in Europe to discredit Ukraine. Le Monde. https://bit.ly/47df6v1
León, J.L. (1993). Persuasión de masas. Deusto.
Lesaca, J. (2017). Armas de seducción masiva: La factoría audiovisual de Estado Islámico para fascinar a la generación millennial. Península.
Lever, R. (2016, 22 noviembre). Breitbart, el polémico portal extremista que ayudó al magnate, se expande. La Nación. https://acortar.link/TMOFT0
Levien, S.J. (2024, 11 julio). ¿Qué es el proyecto 2025 y por qué Trump se distancia de él?. The New York Times. https://acortar.link/2pdJoQ
Lewsey, F. (2025, diciembre 11). Price of a bot army revealed across hundreds of online platforms. University of Cambridge. https://acortar.link/TSHrLF
Linvill, D., & Warren, P. (2024) What's Hiding Under the Kilt? Iranian Trolls for Scottish Independence. Media Forensics Hub Reports, 6. https://open.clemson.edu/mfh_reports/6
Lippmann, W. (1922). Public Opinion. Macmillan.
Lipscomb, D. (2025, 3 enero). How do employers legally monitor employees' social media? Ferretly. https://acortar.link/4aGKEv
López Carrión, A.E., & Llorca-Abad, G. (2025). Desinformación durante la crisis producida por la DANA de 2024 en España: análisis, características, tipologías y desmentidos. Revista Mediterránea De Comunicación, 16(2), e29303. https://doi.org/10.14198/MEDCOM.29303
López-Fonseca, Ó. (2025, 12 septiembre). La policía busca a un profesor por facilitar a un grupo prorruso información para lanzar ciberataques contra España. El País. https://acortar.link/wL8ogR
López Gómez, S., Mendieta Díaz, G., & Micó Faus, J. S. (2021). Tendencias online de la propaganda yihadista. Análisis del caso español (Documento de Opinión No. 85/2021). Instituto Español de Estudios Estratégicos (IEEE). https://bit.ly/43npM83
MacMillan Center Yale (2022, 6 agosto). Paul Brass' scholarship on India's religious, linguistic politics made invaluable contributions. MacMillan Center. https://bit.ly/47H0iUi

Madrigal, M. (2021, 14 agosto). Médicos por la Verdad: la desinformación convertida en negocio en la web. Newtral. https://bit.ly/4qOg20E

Maestre, A. (2025, 1 diciembre). La célula neonazi española The Base planeaba atentar de manera inminente emulando el ataque de Christchurch, Nueva Zelanda. La Sexta. https://acortar.link/mYCAeY

Maldita. (2024, 23 octubre). Consecuencias de la desinformación: insultos, amenazas y campañas de odio contra los servicios de meteorología en España y en el mundo. Maldita.es. https://bit.ly/3LoWFLg

Maldita. (2025, 19 marzo). 'LLM grooming' y cómo esta técnica busca manipular las IA para desinformar. Maldita.es. https://bit.ly/3Jz2rcP

Maldita. (2025a, 30 mayo). Pravda: un medio de desinformación que quiere expandirse en España. https://acortar.link/nlfFny

Maldita. (2025b, 10 julio). Canales de Telegram de Pravda: la maquinaria de desinformación en español. https://acortar.link/xXYxWY

Marshall, A.R.C. & Tanfani, J. (2022, 22 agosto). New breed of video sites thrives on misinformation and hate. Reuters. https://tinyurl.com/3dfrj99w

Martín, A.L. (2022, 22 marzo). Odysee, la plataforma estadounidense en la que RT y Sputnik siguen emitiendo. El Español. https://tinyurl.com/5n977rs2

Martínez Ron, A. (2023). ¿Puede Facebook cambiar unas elecciones? Un macroestudio dice que el algoritmo no altera las opiniones políticas. El Diario. https://acortar.link/wNg7HK

Maurice, K.S. (2024, 28 agosto). Combat misinformation for a stable economy. HKU Business School. https://acortar.link/5LEAlj

Matz, S.C., Menges, J.I., Stillwell, D.J., Schwartz, H.A. (2019) Predicting individual-level income from Facebook profiles. PLoS ONE 14(3): e0214369. https://doi.org/10.1371/journal.pone.0214369

Méndez, B. (2025, 25 octubre). Hackers prorrusos vuelven a poner el foco en España. La Razón. https://acortar.link/MXiJN0

Mejova, Y., Capozzi, A., Monti, C., & De Francisci Morales, G. (2025). Narratives of war: Ukrainian memetic warfare on Twitter. Proceedings of the ACM on Human-Computer Interaction, 9(2), Article CSCW139. https://doi.org/10.1145/3711037

Meta (2019a, 13 enero). Removing Coordinated Inauthentic Behavior From Iran. https://bit.ly/4m6bcIB

Meta (2019c, 6 mayo). More CIB from Russia. https://acortar.link/Gjgh7R

Meta (2019b, 20 septiembre). Removing Coordinated Inauthentic Behavior in Spain. https://bit.ly/45X0KhF

Meta (2020, 24 septiembre). Removing Coordinated Inauthentic Behavior. https://bit.ly/4mWKXW5

Meta (2020, 15 diciembre). Removing Coordinated Inauthentic Behavior from France and Russia. https://acortar.link/p0yGbO

Meta (2021, 12 enero). December 2020 Coordinated Inauthentic Behavior Report. https://bit.ly/45XBJ61

Mettler, T. (2024). The connected workplace: Characteristics and social consequences of work surveillance in the age of datification, sensorization, and artificial intelligence. Journal of Information Technology, 39(3), 547-567. https://journals.sagepub.com/doi/10.1177/02683962231202535

Milgram, S. (1967). The small world problem. Psychological Today, 2(1), 60-67. https://acortar.link/2Zbc6a

Milosevich-Juaristi, M. (2020). ¿Por qué hay que analizar y comprender las campañas de desinformación de China y Rusia sobre el COVID-19? Real Instituto Elcano. https://bit.ly/4hQ3a5Y

Milosevich-Juaristi, M. (2021). La "combinación", instrumento de la guerra de la información de Rusia en Cataluña. Real Instituto Elcano. https://bit.ly/43rS950

Ministère de L'Europe et des Affaires Étrangères (2025). Face aux manipulations d lìnformation. http://bit.ly/46ouumI

MITRE Corporation. (n.d.). MITRE ATT&CK. https://attack.mitre.org/

Moncrieff, M., Kilibarda, P., & Gaggioli, G. (2024). Social network analysis and counterterrorism: A double-edged sword for international humanitarian law. Journal of Conflict and Security Law, 29(1), 165–183.

Monitor Disinfo (2023). Irinamar Z: Análisis de un canal de propaganda rusa en Telegram. https://acortar.link/LdWb6a

Monitor Disinfo (2025). Fin de partida para NoName05716: Europa desmantela una pieza clave de la ciberguerra prorrusa. https://acortar.link/X5hYV6

Mottareale-Calvanese, D., Arce-García, S., & Said-Hung, E. (2025). How does hatred spread in Italy? Mario Draghi's dimission case. Social Sciences & Humanities Open, 12, 102148. https://doi.org/10.1016/j.ssaho.2025.102148

Munk, T. (2025). Digital defiance: Memetic warfare and civic resistance. European Journal of Crime Policy and Research, 31, 501–528. https://doi.org/10.1007/s10610-025-09613-4

Myers, D.G. (2009). Psicologia social. McGraw-Hill.

Naelle-Nemann, E. (1995). La espiral del silencio. Opinión pública: nuestra piel social, Paidós.

Naked Security (2020, 13 enero) Facebook prohíbe los deepfakes, pero no los cheapfakes o shallowfakes. https://bit.ly/3VCCIT6

National Geographic España (2022, 15 junio). Seis grados y las redes sociales: la teoría que conectó el mundo. National Geographic. http://bit.ly/4nyw8sr

Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical Preferences Predict Personality: Evidence From Active Listening and Facebook Likes. Psychological science, 29(7), 1145–1158. https://doi.org/10.1177/0956797618761659

Newtral (2025, 1 octubre). 'Invadidos', una web antiinmigración que usa falsos "periodistas" con fotos rusas para monetizar el odio. https://bit.ly/47l4OsV

Núñez, J. Á. (2025, 14 febrero). Los que "debemos pagar por ello": cuando fui amenazado por informar de la DANA. El País. https://bit.ly/47o5fT8

Nygren, T., Frau-Meigs, D., Corbu, N., & Rossi, R. (2022). Teachers' views on disinformation and media literacy supported by a tool designed for professional fact-checkers: Perspectives from France, Romania, Spain, and Sweden. SN Social Sciences, 2(40). https://doi.org/10.1007/s43545-022-00340-9

Liang, C.S. (2015). Cyber Jihad: Understanding and countering Islamic State propaganda. The Geneva Centre for Security Policy, Policy Paper. https://acortar.link/ZokkNG

OCCRP. (2023, 17 febrero). Eliminalia: A Reputation Laundromat for Criminals. OCCRP. https://bit.ly/3JGMD7H

Olari, V. (2025, mayo 16). From Bucharest to Chisinau: How pro-Kremlin networks are shaping Romania's 2025 election. Digital Forensic Research Lab (DFRLab). https://acortar.link/zBJ0hP

Ollero, D.J. (2019, 7 agosto). Así es 8Chan, la web proscrita en la que los asesinos de masas cuelgan sus manifiestos. El Mundo. https://acortar.link/s6naHq

Olmo, J.M. (2023, 3 febrero). Así es Rafapal, el jefe de la conspiración en España: "Hay clones. ¿Quién los controla?" El Confidencial. https://acortar.link/oIMLG1

Ong, J.C., & Cabañes, J.V.A. (2019). Politics and Profit in the Fake News Factory: Four Work Models of Political Trolling in the Philippines. NATO Strategic Communications Centre of Excellence. https://acortar.link/BD868T

Ong, J.C., Tapsell, R., & Curato, N. (2019) Tracking digital disinformation in the 2019 Philippine Midterm Election. New Mandala. https://www.newmandala.org/disinformation

Orr Bueno, C. (2025, 24 noviembre). X just accidentally exposed a vast misinformation campaign. Weaponized Spaces. https://acortar.link/7Do3PJ

Pamment, J., & et al. (2019). Hybrid Threats: Disinformation in Sweden NATO Strategic Communications Centre of Excellence. https://bit.ly/4qDCzwP

Pariser, E. (2012). The filter bubble: what internet is hiding from you. Penguin Books.

Park, G., Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D.J., Ungar, L.H., & Seligman, M.E. (2015). Automatic personality assessment through social media language. Journal of personality and social psychology, 108(6), 934–952. https://doi.org/10.1037/pspp0000020

Patrikarakos, D. (2017). War in 140 characters: How social media is reshaping conflict in the twenty-first century. Basic Books.

Paulo, D.A. (2022, 4 septiembre). Paid troll army: The influence of hired social media influencers in Philippines elections. Channel News Asia. https://acortar.link/pRGxov

Paz, D. (2020). Esa noticia. Blog de Daniel Paz. https://danielpaz.com.ar/blog/2020/05/esa-noticia/

Peinado, F., Palomo, E. & Galán, J. (2018, 21 marzo). Las redes rotas de la campaña electoral en México. ElPaís.com. https://bit.ly/43ryGkX

Pérez García, Á., Suárez Perdomo, A., López Martínez, A., & Martínez Fernández, G. (2024). Los adolescentes y la construcción del pensamiento crítico para la gestión de los retos y las noticias falsas en las redes sociales. Aloma: Revista De Psicologia, Ciències De l'Educació I De l'Esport, 42(1), 59–67. https://doi.org/10.51698/aloma.2024.42.1.59-67

Pérez Triana, J.M. (2025, 29 julio). "Africa Corps" releva al grupo Wagner en el Sahel tras su disolución: "La misión ha concluido". Elconfidencial.com. https://acortar.link/ep2Svt

Pichi, A. (2024, 14 noviembre). The Onion buys Alex Jones' Infowars at bankruptcy auction. Here are the details. CBS News. https://acortar.link/8wahrr

Pinkola Estés, C. (2012, 5 enero). Newt Gingrich: And His List of Words. The Moderate Voice. https://acortar.link/ZgMq6s

Planas Bou, C. (2021, 04 marzo). Qué es I3 Ventures, la empresa en el núcleo del 'Barçagate'. El Periodico. https://bit.ly/3KqJNzC

Plutchik, R. (1980). Emotion: A Psychoevolutionary Synthesis. Harper & Row.

Poliakoff, S. (2025). Trolls behind the mask of journalists: How Yevgeny Prigozhin's Patriot Media Group was organized. Problems of Post-Communism. https://doi.org/10.1080/10758216.2024.2438336

Pomerantsev, P. (2024). How to win an information war: The propagandist who outwitted Hitler. Farrar, Straus and Giroux.

Pontijas, J.M. (2020). Control reflexivo: mucho más que desinformación a la rusa. Instituto Español de Estudios Estratégicos. https://acortar.link/SHgNVe

Posetti, J., Maynard, D., Bontcheva, K., & Shabbir, N. (2024). Carole Cadwalladr: The networked gaslighting of a high-impact investigative reporter. International Center for Journalists. https://acortar.link/eWcAPV

Pozas, A. (2023, 20 enero). La Justicia condena al agitador ultra Alvise Pérez a indemnizar a la periodista Ana Pastor por vulnerar su honor. Eldiario.es. https://tinyurl.com/3hu7fcn9

ProBlock (2021). ProBlock: a novel approach for fake news detection. https://acortar.link/U06Zfl

Psaledakis, D. (2025, 16 abril). US State Department closing office aimed at countering foreign disinformation. Reuters. https://acortar.link/SHrINr

Psychological Defense Agency (2025). Cuando nuestros pensamientos nos llevan por mal camino. http://bit.ly/47Nnnq5

Pugliese, D. (2020, 13 octubre). Canadian military spent more than $1 million on controversial propaganda training linked to Cambridge Analytica parent firm. Ottawa Citizen. https://bit.ly/38LBjVH

QuantumSEC (2024, abril). Palantir Technologies: Anatomía de un Imperio de Datos. https://bit.ly/4qTIb6v

Quénel, N. (2023). Allô, Paris? Ici Moscou. Plongée au cœur de la guerre de l'information. Denoël.

Quiñones de la Iglesia, F.J. (2021). Desinformación y subversión (2.0): las técnicas de la Guerra Fría reaparecen en el dominio informativo del siglo XXI. Documento Marco, 12/2021.

Ram, Y., Wiener, A., & Ring, E. (2025). Disinformation resilience: Israel stakeholder analysis report. Israel Internet Association. https://acortar.link/ijJRUZ

Recuero, R., Guazina, L.S., & Araújo, B. (2025). State-sponsored disinformation in Brazil: Distrust and delegitimisation of the electoral system through the use of political authority Facebook accounts. En M. Echeverría, S. G. Santamaría, & D. C. Hallin (Eds.), State-Sponsored Disinformation Around the Globe: How Politicians Deceive their Citizens (pp. 139–156). Routledge. https://doi.org/10.4324/9781032632940-12

Redacción LR. (2021, 18 mayo). España: normativa provoca discriminación de género en el ajedrez. La República. https://bit.ly/4nIDGch

Redfish (2017, 16 noviembre). Catalan Independence: Fighting Franco's Ghost [Trailer] [Video]. Internet Archive. https://acortar.link/zD2K5g

Reporteros Sin Fronteras (2017, marzo). Las "fake news", un pretexto de los predadores de la libertad de prensa para censurar. https://bit.ly/3XprXEq

Rey García, P., Rivas Nieto, P., & Sánchez Alonso, Ó. (2017). Propaganda, radicalismo y terrorismo: la imagen del Daesh. Estudios sobre el Mensaje Periodístico, 23(1), 209-221. https://doi.org/10.5209/ESMP.55592

Rid, T. (2020). Desinformación y guerra política: Historia de un siglo de falsificaciones y engaños (Trad. de A. Martínez). Debate.

Riedl, M.J., Strover, S., Cao, T., Choi, J.R., Limov, B., & Schnell, M. (2021). Reverse-engineering political protest: the Russian Internet Research Agency in the Heart of Texas. Information, Communication & Society, 25(15), 2299–2316. https://doi.org/10.1080/1369118X.2021.1934066

Ríos Gutiérrez, J.A. (2024). Epílogo: La inteligencia artificial y la desinformación. Comunicación Científica, (244), 79–89.

Rodríguez Fernández, L. (2021). Propaganda digital: Comunicación en tiempos de desinformación. Editorial UOC.

Rodríguez Fernández, L. & et alters (2023). Mapa de las capacidades de investigación en materia de desinformación en las universidades y centros de investigación españoles. En Foro contra las campañas de desinformación en el ámbito de la seguridad nacional. Trabajos 2023 (pp. 116-167). Presidencia del Gobierno, Departamento de Seguridad Nacional. https://acortar.link/VvtTft

Rodríguez-Fernández, L. (2024). La industria de la desinformación: aproximación a sus orígenes y evolución. En J. M. Aguado, M. V. de Haro de San Mateo, Á. Gómez de Ágreda y M. Pérez-Escolar (Eds.), Desinformación y defensa: Conflictos híbridos, entorno cognitivo y operaciones de influencia (cap. 5). Dykinson.

Rodríguez-Fernández, L., González-Fernández, S., & Arce-García, S. (2025). Desinformación sobre procesos electorales en España: El caso de la plataforma Elecciones Transparentes. Revista Mediterránea de Comunicación, 16(2), e28554. https://doi.org/10.14198/MEDCOM.28554

Rodrixoc (2019). Arriando la bandera española de Cuba en 1898. http://bit.ly/4oZBZsx

Rosen, G. (2022, 17 mayo). Community Standards enforcement report, first quarter 2022. Meta. https://bit.ly/42hxEqU

RTVE (2020, 9 julio). Estrasburgo condena a Rusia por el derribo del vuelo MH17 por violaciones de derechos humanos en Ucrania. https://acortar.link/3EUyl9

RTVE (2023). El proyecto IVERES de Inteligencia Artificial presenta los primeros resultados en la lucha contra la desinformación. https://acortar.link/UcuKyC

RTVE (2024, 30 abril). La Comisión Europea investiga a Meta por considerar que no combate la desinformación ante las elecciones europeas. https://bit.ly/4pdaewZ

Russian Presidential Library. (n.d.). The history of the Russian state: The role of information in the modern world. https://www.prlib.ru/history/619826

Sánchez, J. (2019, 7 marzo). Trump y la mujer que se enfrenta a los trolls rusos: La periodista finlandesa Jessikka Aro ha sido objeto de amenazas y ataques tras investigar la manipulación en redes sociales. La Vanguardia. https://acortar.link/Fka87V

Sánchez-Vallejo, M. A. (2025, 14 octubre). El Supremo de EE UU rechaza un recurso de Alex Jones para evitar pagar 1.400 millones de dólares a las familias de Sandy Hook. El País. https://acortar.link/Qc3nTC

Sanchis, A., & Alamillos, A. (2025, 14 julio). Torre-Pacheco, la bomba de relojería de los bulos prorrusos en redes como Telegram. El Confidencial. https://acortar.link/I3AH9G

Sanger, D.E., Barnes, J.E., & Conger, K. (2022, 1 marzo). Los tanques entraron a Ucrania y también los programas de 'malware'. Entonces Microsoft se involucró en la guerra. The New York Times. https://bit.ly/3Lzj9cy

Sanger, D.E., Barnes, J.E., & Goldman, A. (2020, 10 julio). Trump claims credit for 2018 cyberattack on Russia. The New York Times. https://acortar.link/3F7drH

Santini, R.M., & Salles, D. (2025). Brazil's Far-Right 'Media Literacy' Campaign: Delegitimizing the Press for Profit. In E. L. Briant & V. Bakir (Eds.), Routledge Handbook of the Influence Industry. Routledge. https://doi.org/10.4324/9781003256878

Sauer, P. (2023, 8 julio). Putin se apodera del imperio empresarial del jefe de Wagner y las granjas de trolls. El Diario. https://acortar.link/dLNF76

Schlesinger, S., & Kinzer, S. (1982). Bitter fruit: The story of the American coup in Guatemala. Harvard University Press.

Schoch, D., Keller, F.B., Stier, S. et al. (2022). Coordination patterns reveal online political astroturfing across the world. Sci Rep 12, 4572. https://doi.org/10.1038/s41598-022-08404-9

Secrétariat général de la défense et de la sécurité nationale-SGDSN. (2025a, febrero). Viginum : Rapport public sur les risques liés aux élections en Roumanie pour la France. https://acortar.link/Gdx6gu

Secrétariat général de la défense et de la sécutité nationale-SGDSN. (2025, junio). African Initiative: from public diplomacy covert influence operations. Viginum. https://acortar.link/s2mYuq

Semple, K., & Ahmed, A.S. (2018, 24 junio). El PRI y el fantasma del fraude electoral en México. The New York Times. https://bit.ly/4p0fgM4

Sewell Jr, W. H. (1996). Three Temporalities: Toward an Eventful Sociology. The Historic Turn in the Human Sciences, 245-280.

Shiundu, A., & Jiménez, A. (2022, 22 febrero). New evidence of thriving 'disinformation industry' on Twitter is worrying as Kenya gears for elections in August 2022. Africa Check. https://bit.ly/3zH2Thw

Shultz, B. (2025). Ecoboost: an environmentally focused Russian influence network targets Norway. Amercian Sunlight Project and Bellona. https://bit.ly/4niLdys

Siegelman, W. (2018, 24 junio). Social media and influence: Companies related to the Trump-Russia story. News Tracs. https://acortar.link/UBG7d5

Siegelman, W. (2022, 14 agosto). Trump Media & Technology Group, Rumble Inc. y Cosmic Development. NewsTracs.com. https://tinyurl.com/4xkwe2jh

Siegelman, W. (2024a, 15 noviembre). Cambridge Analytica, Emerdata y un exdirector vinculado al proyecto de criptomonedas de Trump. News Tracs. https://acortar.link/BXBubC

Siegelman, W. (2024b, 8 julio). Update on Cambridge Analytica funder Rebekah Mercer who runs Emerdata, co-founded 1789 Capital with Omeed Malik and is a key trustee of the Heritage Foundation which launched Project 2025. https://acortar.link/0u7Tf9

Siegelman, W. (2024c, 15 noviembre). Update on Cambridge Analytica, its spawn Emerdata Limited, and a former director who now advises a Trump-supported crypto project. NewsTracs. https://acortar.link/BXBubC

SimplVest (2025, 4 abril). The Rise of Yahoo Schools in Nigeria: How Young Men Are Groomed for Cybercrime. https://acortar.link/noDkx4

Singer, F.M. (2023, 20 febrero). No son periodistas, son avatares: el chavismo impulsa propaganda hecha con inteligencia artificial. ElPaís.com. https://acortar.link/CZR64l

Skibinski, M. (n.d.). Brands send billions to misinformation websites: Newsguard & Comscore report. NewsGuard. https://acortar.link/qG9oKw

Snowden, E. (2019). Vigilancia permanente. Editorial Planeta.

Sobchuk, M. (2025, 13 junio). To understand the nature of modern chinese influence operations, study Russia first. Cyfluence Research Center-CRC. https://acortar.link/13bKlT

Soldatov, A., & Borogan, I. (2015). The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries. PublicAffairs

South China Morning Post (2009, 13 enero). Beijing in 45b yuan global media drive. https://acortar.link/HVRgbl

Sprinforma (2024, 19 mayo). Revelan campaña de red de bots y trolls para propagar el rumor de que EEUU tiene una "lista negra" de políticos mexicanos. https://bit.ly/47Vgq4C

Stajner, S., & Yenikent, S. (2021). Why Is MBTI Personality Detection from Texts a Difficult Task?. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 3580–3589, Online. Association for Computational Linguistics.

Stella, M., Ferrara, E., & de Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences of the United States of America, 115(49), 12435-12440. https://doi.org/10.1073/pnas.1803470115

Stengel, R. (2019). Information Wars: How We Lost the Global Battle Against Disinformation and What We Can Do About It. PublicAffairs.

Stolze, M. (2022). Information laundering via Baltnews on Telegram: How Russian state-sponsored media evade sanctions and narrate the war. NATO Strategic Communications Centre of Excellence. https://bit.ly/48U00f7

Stracqualursi, V. (2020, 29 septiembre). Trump's 2016 campaign used voter deterrence techniques and fears of fraud to suppress turnout. CNN. https://acortar.link/Y6HqSr

StratComCOE & HybridCOE (2021). DISARM report: A comprehensive analysis of disinformation and hybrid threats. http://bit.ly/47Vd0As

Strategic Analysis (2023). Dissemination of pro-Russian narratives via media in North Macedonia. Strategic Analysis Think-Tank. https://acortar.link/5oGpw2

Suárez-Ruiz, E. J., & González Galli, L. (2022). Alfabetización digital como ética preventiva: educación metacognitiva para el contexto mediático post COVID-19. AdComunica, (23), 119–140. https://doi.org/10.6035/adcomunica.6201

Suau, J., & Puertas-Graell, D. (2023). Disinformation narratives in Spain: reach, impact and spreading patterns. Profesional de la información, 32(5), e320508. https://doi.org/10.3145/epi.2023.sep.08

Swissinfo (2023, 3 noviembre). Meta eliminó miles de contenidos de desinformación en España en torno a las elecciones de la UE. https://acortar.link/Ph3sUf

Tadaweb, & Charon, P. (2025, octubre). Baybridge, anatomy of a chineses information influence ecosystem. IRSEM, Institute de Recherche Stratégique de l'École Militaire. Ministère des Armées et des Anciens Combattants. https://acortar.link/rRe0rv

Tajfel, H., Billig, M.G., & Bundy, R.P. (1971). Social categorization and intergroup behavior. European Journal of Social Psychology, 1, 149–178.

Tchakhotine, S. (1992). Le Viol des foules par la propagande politique. Gallimard.
Terrero Carrobles, J.J. (2025, 17 noviembre). El cambio de estrategia del espionaje ruso en Europa. Ministerio de Defensa - CESEDEN. https://acortar.link/x8zUEu
The American Leader (1990). Language: A Key Mechanism of Control. https://acortar.link/B6bWsb
The Graphika Team (2025, 29 enero). Chinese State Influence. https://acortar.link/fi8IhT
The Guardian (2017, 10 enero). El Movimiento 5 Estrellas se divorcia de UKIP pero los liberales lo plantan en el altar. elDiario.es. https://bit.ly/3XoUgD1
The Influence Explorer (2025). Explorer. https://influenceindustry.org/en/explorer/
The Influence Industry Project (2022). The Influence Industry Project. https://influenceindustry.org/
Thomas, T. (2004). Russia's Reflexive Control Theory and the Military. The Journal of Slavic Military Studies, 17(2), 237–256. https://doi.org/10.1080/13518040490450529
Thomas, J., & França, T. (2025, 18 abril). Euroverify: Qué países europeos están más expuestos a la desinformación rusa. https://acortar.link/n2M6cd
Traufetter, G., Hesse, U., & Jung, A. (2023, 7 mayo). Moskaus Taschengeld-Agenten – es fehlt nur die Reichweite. Süddeutsche Zeitung. https://bit.ly/42ZWv2Y
TruthAfrica (2025, 18 noviembre). A battle for hearts and minds: How Russian propaganda takes over Africa. EU vs Disinfo. https://acortar.link/sdNMMH
Tsygankov, A.P., & Tsygankov, P.A. (2021). Constructing National Values: The Nationally Distinctive Turn in Russian IR Theory and Foreign Policy. Foreign Policy Analysis, 17(4), orab022. https://doi.org/10.1093/fpa/orab022
Tye, L. (1998). The father of spin: Edward L. Bernays and the birth of public relations. Crown Publishers.
Universidad de Palermo (2023). Comunicación, cultura y política en la era de la desinformación, 18(2). https://acortar.link/OL8mjE
University of Cambridge Social Decision-Making Lab. (2025). Cambridge Online Trust and Safety Index. https://cotsi.org/
University of New South Wales (2023). Understanding mass influence: A case study of Russia's Internet Research Agency (IRA). UNSW Canberra. https://acortar.link/9oXim2
Uyheng, J., & Carley, K. M. (2021). Characterizing network dynamics of online hate communities around the COVID-19 pandemic. Applied Network Science, 6(1), 1–21. https://doi.org/10.1007/S41109-021-00362-X
Valencia, J. (2023). México y la guerra de información en LATAM. Monitor.disinfo. https://acortar.link/wWbWh8
Vargas Pasaye, R.G. (2022). La Conferencia mañanera de AMLO y la cultura de la cancelación. Razón Y Palabra, 26(113). https://doi.org/10.26807/rp.v27i113.1871

Varios autores (2025). Mediated Communication, Public Opinion and Society Section. IAMCR 2025 Congress Singapore. https://bit.ly/4qfRy05

Vecchione, M, Schoen, H., González Castro, J.L., Cieciuch, J., Pavlopoulos, V., & Caprara G.V. (2011). Personality correlates of party preference: The Big Five in five big European countries. Personality and Individual Differences, 51(6), 737-742, https://doi.org/10.1016/j.paid.2011.06.015

Verificado (2020, 19 mayo). Influencers compartieron información falsa antivacunas. Las consecuencias podrían ser más graves de lo imaginado. Verificado. https://bit.ly/3JxWNrq

Vidgen, B. (2019). Tweeting Islamophobia. Doctoral Thesis, Oxford University Research Archive, University of Oxford. https://bit.ly/4p4agGN

Viginum (2024, junio). Matriochka. Une campagne prorruse ciblant les médias et la communauté des fact-checkers. Rapport technique. https://acortar.link/NELUIl

Vila Márquez, F., & Arce García, S. (2019). Fake News y difusión en Twitter: el caso de Curro, el perro "condenado". Historia Y Comunicación Social, 24(2), 485-503. https://doi.org/10.5209/hics.66292

Vilmer, J-B. J., Escorcia, A., Guillaume, M., & Herrera, J. (2018). Les manipulations de l'information. Un défi pour nos démocraties. Centre d'analyse, de prévision et de stratégie (CAPS) du Ministère de l'Europe et des Affaires étrangères et de l'Institut de recherche stratégique de l'École militaire (IRSEM) du Ministère des Armées. https://acortar.link/bpsaQX

Walicek, T. (2023, 6 febrero). The $340 million anti-labor consulting industry is behind contemporary union-busting. Truthout. https://acortar.link/QfRjSr

Wallace, J., Goldsmith-Pinkham, P., & Schwartz, J. L. (2023). Excess Death Rates for Republican and Democratic Registered Voters in Florida and Ohio During the COVID-19 Pandemic. JAMA internal medicine, 183(9), 916–923. https://doi.org/10.1001/jamainternmed.2023.1154

Wendling, M. (2016, 21 noviembre). ¿Preferirías que tu hija fuera feminista o tuviera cáncer? BBC Mundo. https://acortar.link/8AdxKK

Wesolowski, K. (2023, 15 febrero). Faktencheck: Was wir über Israels bezahlte Propaganda-Werbekampagne wissen. Deutsche Welle. https://acortar.link/3lDRzB

Williams, M. (2021) The Science of Hate. Faber & Faber.

Wylie, C. (2020). Mindf*ck Cambridge Analytica. La trama para desestabilizar el mundo. Roca Editorial de Libros, S. L.

Wywiał, P. (2023). Joke and meme as weapons in the Ukrainian war. Institute for Security and Informatics. https://doi.org/10.34739/dsd/2023/02.03

Wolf, M. (1987). La investigación de la comunicación de masas. Barcelona: Paidós.

Yair, A., & Perlov, O. (2024). Artificial intelligence in the service of Israel's public diplomacy. Institute for National Security Studies (INSS). https://acortar.link/7Dk8Hk

Yasur, N., Ram, Y., & Ring, E. (2025). Ballistic Fakes: Disinformation and Fact-Checking Efforts during the Israel–Iran War. Israel Internet Association. https://acortar.link/Cq2jIC

Zeitzoff, T. (2016). Does social media influence conflict? Evidence from the 2012 Gaza conflict. Journal of Conflict Resolution, 62(1), 29–63. https://www.jstor.org/stable/48597288

24 Horas Puebla. (2017, 1 enero). Grupo que promovió SaqueaUnWalmart es español. https://acortar.link/huODup

@DouglasMun (2024, 22 marzo). Evolution of click farm fraud.[Post]. X. https://bit.ly/4ohAqph

La mentira no es nueva. Lo nuevo es la industria que la fabrica.

Desde que Ramsés II mandó esculpir en piedra en el antiguo Egipto una victoria que nunca ocurrió, la humanidad ha librado guerras antes de que los ejércitos se movieran. Hoy, esa batalla se libra en la palma de tu mano.

La anatomía de la desinformación. Redes, bots y odio disecciona con bisturí académico y mirada periodística la maquinaria global que fabrica mentiras, amplifica el odio y polariza sociedades enteras. Un recorrido que arranca en Edward Bernays (el hombre que convenció a las mujeres de fumar en nombre de la libertad) y atraviesa Cambridge Analytica, los foros de 4chan, las granjas de trolls del Kremlin, Qanon y la guerra de Ucrania, hasta llegar al presente, donde los algoritmos conocen tus miedos mejor que tú mismo.

Pero este libro no se limita a contar qué ocurre. Muestra cómo se construye una campaña de desinformación y odio paso a paso: cómo se planifica, cómo segmenta a sus víctimas, qué emociones activa y por qué funciona. Porque para desarmar una trampa, primero hay que entender cómo está tendida.

Una obra para quien quiera comprender el mundo en el que vive, y para quienes se niegan a ser manipulados.

EDICIONES Universidad de Valladolid

COLECCIÓN comunicación