Guía del alumno

3.8 · Formas de manipulación organizada
Módulo 3: Modelos, mecanismos y redes sociales
Capítulo 9: Operaciones de influencia en redes sociales · L26 de 38

G1 · Propósito y recorrido

Una operación puede engañar sobre el origen de sus voces y sobre la independencia del apoyo que recibe, incluso cuando utiliza contenidos verdaderos. Estudiar esa organización exige separar identidad, control, coordinación y patrocinio. Esta guía enseña a clasificar conductas y a leer investigaciones sin convertir una sospecha en una certeza.

Resultados de aprendizaje

- RA1. Distinguir anonimato, identidad ficticia y suplantación; diferenciar automatización, gestión humana e intervención híbrida.
- RA2. Explicar cómo la falsa independencia y la amplificación artificial pueden construir apariencia de consenso; separar contenido, comportamiento y patrocinio.
- RA3. Interpretar un informe de retirada de redes y un experimento de interacción artificial, identificando sus métodos, hallazgos y límites.
- RA4. Formular un diagnóstico justificado sobre un expediente y señalar qué pruebas faltan para sostener otras conclusiones.

Cómo estudiar esta lección

1. Sigue la presentación y su explicación. 2. Trabaja G2–G7, comparando definiciones y casos. 3. Resuelve G8 y G9 sin consultar G10. 4. Contrasta tus razones con las soluciones. 5. Realiza el cuestionario limpio. Reserva unos 45–60 minutos para el estudio y las actividades, además del vídeo.

Continuidad con el curso

La 3.7 explicó el entorno digital y la circulación. La 3.8 examina organización y engaño en la participación. La 3.9, «Efectos medidos y límites de la evidencia», desarrollará la medición de efectos; la 3.10, la identificación y resistencia. Esta lección no inicia ni cierra capítulo o módulo. Las referencias [F1]–[F5] remiten a las fuentes completas en G11.

G2 · Identidad: qué afirma ser una cuenta

La identidad pública es una representación. Puede describir fielmente al emisor, reservar una parte de su identidad o construir una falsa apariencia. La ausencia de un nombre civil no demuestra manipulación. Debe examinarse qué afirma el perfil y si esa afirmación engaña sobre quién participa.

Situación	Rasgo decisivo	Ejemplo didáctico
Anonimato o seudónimo	El nombre no se revela o se sustituye; no implica apropiarse de otra identidad.	Una autora firma «Bitácora» sin fingir ser otra persona.
Persona ficticia engañosa	Se inventa una biografía y se presenta como persona real.	Una empresa crea «una vecina» inexistente para aparentar apoyo local.
Suplantación	Se hace pasar por una persona o institución concreta sin autorización.	Un perfil afirma ser la cuenta oficial de una universidad que no lo gestiona.
Parodia identificada	La representación se presenta de forma reconocible como parodia.	Un perfil indica expresamente que no es la cuenta de la institución imitada.

Cómo reconocer la diferencia

Pregunta qué promesa de identidad ofrece el perfil: ¿una voz reservada, una biografía supuestamente real o la representación de alguien concreto? Contrasta esa promesa con pruebas pertinentes. Un registro de administración puede vincular cuentas a un operador; una declaración institucional puede desmentir una falsa representación. Una fotografía genérica, por sí sola, no demuestra ninguna de las dos cosas.

Cuentas intervenidas y continuidad aparente

Una cuenta auténtica puede ser robada o intervenida. Su historial y sus seguidores no garantizan que siga bajo el mismo control. Oxford agrupa cuentas intervenidas, robadas y de suplantación dentro de una categoría de su inventario; eso no convierte sus mecanismos en idénticos. [F1, p. 11/PDF 13]

Comprobación breve: si alguien escribe bajo seudónimo y comparte opiniones impopulares, ¿qué prueba existe de suplantación? Ninguna de esas dos características la acredita. Hace falta demostrar la falsa representación, no juzgar la opinión.

G3 · Control, automatización y apoyo aparente

Quién ejecuta las acciones

Un bot ejecuta acciones mediante software automatizado. Una cuenta administrada por personas depende de decisiones e intervenciones humanas; puede usar programación auxiliar. Un sistema híbrido combina ambas formas. No son categorías de verdad y mentira. Un bot declarado puede ser legítimo; un operador humano puede controlar identidades engañosas. Tampoco debe confundirse «trol» —conducta provocadora o perturbadora, según el contexto— con «bot», que describe ejecución. [F1, p. 11/PDF 13; F4, pp. 22–24/PDF 40–42]

Apariencia de consenso: mecanismo propuesto

Cuando un mismo centro controla varias voces que se presentan como independientes, sus coincidencias pueden parecer confirmaciones separadas. Si además se apoyan mediante comentarios o reacciones, pueden dar la impresión de un respaldo espontáneo más amplio. El intento de hacer pasar apoyo organizado por movimiento de base espontáneo suele denominarse astroturfing. Aquí usamos el término para ese mecanismo, no para toda campaña coordinada.

El mecanismo pretende que el receptor use señales de popularidad o coincidencia como indicios de aceptación o credibilidad. Eso no prueba que lo consiga. El receptor puede desconfiar, reconocer la operación o ignorarla. Las acciones de la red y la respuesta de la audiencia deben documentarse por separado. El análisis causal de efectos se desarrolla en la 3.9.

Amplificación artificial

Consiste en inflar señales de visibilidad o participación mediante actividad que no representa el interés independiente que aparenta: por ejemplo, interacciones comerciales inauténticas. Una campaña publicitaria identificada o una difusión acordada entre voluntarios no equivale, por ese solo hecho, a tal engaño. El criterio no es simplemente que haya recursos, organización o crecimiento.

Pruebas y límites

Registros de ejecución pueden acreditar automatización; registros de administración pueden acreditar control común. Mucha actividad o publicaciones simultáneas son indicios compatibles con explicaciones alternativas: programación legítima, un acontecimiento urgente o una convocatoria pública. Hay que contrastar esas alternativas antes de afirmar una operación encubierta. Oxford documenta que pueden intervenir cuentas reales y falsas, gestionadas por humanos o mediante automatización. [F1]

G4 · Organización, patrocinio y contenido

Patrocinador, contratista, operador y cuenta son unidades distintas. Quien financia puede contratar a una empresa; la empresa puede dirigir a operadores; estos pueden administrar varios activos. Es un modelo de funciones posibles, no una cadena que deba existir en toda operación. Una persona también puede reunir varias funciones. Oxford identifica organismos públicos, partidos, empresas y colaboradores humanos. [F1, pp. 8–10/PDF 10–12]

Qué hace inauténtica una coordinación

Coordinar significa actuar mediante acuerdos o relaciones organizativas. El comportamiento inauténtico añade engaño sobre identidad, representación o independencia. Una asociación que anuncia su campaña y moviliza miembros reales no equivale a una agencia que hace pasar empleados por vecinos sin vinculación. Tampoco toda colaboración remunerada es inauténtica: la transparencia y la representación importan.

Contenido	Participación transparente	Participación engañosa
Verdadero	Una institución identificada publica un horario correcto.	Perfiles ficticios comparten una noticia cierta y fingen independencia.
Falso o erróneo	Una cuenta identificada publica una cifra equivocada; falta establecer intención.	Una red oculta difunde un dato fabricado y simula apoyo independiente.

Matriz conceptual con ejemplos ficticios. La verdad de una afirmación requiere contrastarla. La autenticidad del comportamiento requiere estudiar quién se presenta, quién controla y qué relaciones oculta. Un error tampoco acredita por sí solo intención de engañar.

Atribuir con precisión

Un vínculo administrativo probado conecta activos y operadores. Un contrato comprobado puede vincular contratista y cliente. Compartir noticias oficiales, usar un idioma o defender una causa no prueba ese encargo. Una atribución de plataforma debe identificarse como tal y conservar su alcance: país de origen, empresa operadora y patrocinador final no significan lo mismo.

Ejemplo público adicional: Facebook informó en 2019 de una retirada de 265 activos y vinculó parte de la actividad a Archimedes Group. Describió cuentas falsas y apariencia de medios locales. Es una atribución de la plataforma; «265 activos» no significa «265 bots» ni mide personas persuadidas. [F5]

G5 · Caso documentado: Unheard Voice

Contexto, actores y propósito

En julio y agosto de 2022, Twitter y Meta retiraron conjuntos parcialmente coincidentes de activos y facilitaron datos a Graphika y Stanford Internet Observatory. El informe del 24 de agosto examinó la actividad encubierta: campañas favorables a intereses occidentales dirigidas a públicos de Oriente Medio, Irán, Afganistán y Asia Central. Los autores infieren ese propósito de los mensajes y patrones; no aportan un mandato del patrocinador. [F2, pp. 2–4/PDF 3–5]

Método y acciones observadas

El equipo combinó contenido, metadatos, perfiles, redes de seguidores e investigación abierta. Separó actividad oficial abierta de campañas encubiertas de origen organizativo no resuelto. Documentó personas ficticias, medios que se presentaban como independientes, imágenes manipuladas, contenidos reciclados y coordinación. El subconjunto de Asia Central difundía contenido en varias plataformas y reutilizaba material de otras fuentes. [F2, pp. 11–15/PDF 12–16]



Figura documental: izquierda, imagen alterada utilizada por una identidad; derecha, foto de Valeria Menendez identificada por los autores. Graphika y SIO, 2022, figura 10, p. 13/PDF 14; recorte de la figura, sin alterar las fotografías. La identidad figuraba como contacto de Intergazeta. No se atribuye participación a la actriz.

Recepción y límites de atribución

La mayoría de las publicaciones examinadas obtuvo pocas interacciones; solo el 19 % de los activos encubiertos identificados superaba mil seguidores. Esto no mide actitudes ni representa todas las operaciones existentes. Twitter indicó como países presuntos de origen Estados Unidos y Gran Bretaña; Meta señaló Estados Unidos. El informe no atribuye públicamente las campañas encubiertas a una entidad concreta. Los vínculos limitados con actividad abierta no resolvieron esa relación. [F2, pp. 2–3/PDF 3–4]

La adquisición inauténtica de seguidores, posiblemente mediante compra, fue una interpretación de los autores basada en patrones de crecimiento y caída, no una transacción comprobada. [F2, p. 12/PDF 13]

G6 · Experimento: interacción artificial en 2024

Qué investigó

Bergmanis-Korāts y Haiduchyk publicaron en noviembre de 2024 un estudio del NATO StratCom COE, con experimento realizado por Trementum Research. Evaluaron capacidades de plataformas para detectar y retirar manipulación comercial. La prueba se realizó en agosto–septiembre de 2024; las conversaciones posteriores con empresas tuvieron lugar en octubre. La publicación declara que no representa opiniones ni políticas de la OTAN o del centro. [F3, p. 2]

Diseño

Los investigadores crearon 44 publicaciones no políticas y compraron interacciones a seis proveedores. Observaron Facebook, Instagram, YouTube, TikTok, VKontakte y X. Midieron entrega, persistencia de cuentas y actividad, entre otros criterios. La compra controlada permitió conocer que las interacciones de la prueba eran comerciales e inauténticas; no fue una encuesta de usuarios ni un experimento sobre intención de voto. [F3, pp. 6–7]

Hallazgos que utilizamos

Unidad observada	Resultado comunicado	Interpretación válida
Interacción artificial	86 % permanecía tras cuatro semanas.	Persistencia de actividad en las condiciones del ensayo.
Cuentas identificadas	Se retiró, en promedio, el 15 % en cuatro semanas.	Retirada de cuentas; no es el complemento del 86 % anterior.
Cuenta frente a actividad	La actividad podía permanecer tras retirar cuentas.	Son resultados y denominadores distintos.

Fuente: F3, pp. 9–10, paginación impresa igual a PDF. El 15 % no permite calcular un «85 % de interacción restante»: cuenta y actividad no son la misma unidad. En las láminas se usa solo el 86 % para evitar mezclar indicadores.

Qué no demuestra

La prueba es acotada, no política, con proveedores seleccionados y una ventana de cuatro semanas. No representa todo el tráfico, todos los países o todas las operaciones. Tampoco demuestra causalmente el efecto de una regulación, aunque los autores discutan esa posibilidad. No midió cuántas personas aceptaron los mensajes ni cuánto cambió su conducta. Los resultados corresponden a 2024 y no son una clasificación actual de las plataformas. [F3, pp. 5–7, 25–26]

Los autores interpretan la pequeña escala como una posible explicación de dificultades de detección. Es una explicación propuesta, no un mecanismo causal aislado por el experimento.

G7 · Comparar investigaciones sin mezclarlas

Fuente	Qué permite conocer	Límite decisivo
Oxford: inventario 2020, publicado en 2021 [F1]	Actores, cuentas y técnicas documentadas en varios países.	Documentación pública desigual; no es un censo ni mide persuasión.
Graphika y SIO, 2022 [F2]	Conductas y contenidos de conjuntos retirados, investigados con datos y fuentes abiertas.	Muestra seleccionada por retiradas; atribución técnica incompleta.
NATO StratCom COE, 2024 [F3]	Entrega y persistencia de interacción artificial conocida en una prueba.	Contexto experimental acotado; no mide opiniones ni representa toda la plataforma.

Coincidencias y diferencias

Las tres investigaciones permiten estudiar organización o participación inauténtica. Oxford amplía el mapa de actores y formas; Graphika y SIO reconstruyen casos; el experimento prueba la persistencia de actividad comercial bajo condiciones conocidas. Sus resultados no son intercambiables. La escasa interacción de muchas cuentas de Unheard Voice no contradice la entrega de interacción comprada en el ensayo: examinan redes, periodos, procedimientos y resultados diferentes.

Cómo se construyó el inventario de Oxford

Bradshaw, Bailey y Howard combinaron análisis de noticias, revisión de literatura y archivos públicos, fichas nacionales y consulta experta. Identificaron cuentas humanas y automatizadas, así como actores estatales, partidistas y comerciales. Su selección deliberada de documentación permite identificar formas observadas. La cobertura mediática, lingüística y documental limita qué casos se conocen. Más casos documentados pueden reflejar tanto actividad como mejor observación; ausencia de registro no equivale a inexistencia. [F1, pp. 6–11/PDF 8–13]

Una ficha de lectura para cualquier informe

- Identifica autor, fecha, fuente de los datos y unidad: cuentas, páginas, publicaciones o personas.
- Separa conductas observadas, interpretaciones y atribuciones de terceros.
- Comprueba el denominador y el periodo de cada cifra.
- Especifica qué recepción se observó y qué efectos no se midieron.

Prueba de comprensión: ¿pueden cien cuentas pertenecer a un operador? Sí. ¿Demuestran cien cuentas cien personas persuadidas? No: ni siquiera permiten identificar por sí solas cien personas distintas.

G8 · Actividad 1: expediente Puerto Claro

Caso enteramente ficticio. Los datos siguientes se consideran comprobados dentro del ejercicio; no describen personas ni organizaciones reales. Tiempo orientativo: 12–15 minutos. Resuelve antes de consultar G10.

Expediente disponible

- D1. La asociación Barrio Abierto publica que doce voluntarios compartirán a las 18:00 una convocatoria vecinal. Sus perfiles identifican su pertenencia a la asociación.
- D2. La oficina municipal publica avisos mediante un programa. El perfil institucional indica que las alertas son automáticas. Los horarios difundidos coinciden con el documento municipal.
- D3. Un registro de administración autenticado vincula veinte perfiles a la agencia Faro. Sus biografías afirman ser vecinos independientes; el expediente confirma que esas personas no existen.
- D4. Registros de trabajo y la revisión de las sesiones acreditan que empleados de Faro redactaron manualmente los comentarios examinados. La orden interna era aumentar la apariencia de respaldo local.
- D5. Esos perfiles compartieron un horario municipal correcto y comentarios de apoyo. No se dispone de contrato con un cliente, de pagos del patrocinador final ni de información sobre recepción fuera de las cuentas de la agencia.
- D6. Una bloguera usa «Brújula» como seudónimo. No se presenta como ninguna persona o institución concreta. No hay otros datos sobre su vinculación con las tres iniciativas.

Tareas

- A1. Clasifica las cuatro situaciones: asociación, oficina, agencia y bloguera. Distingue identidad, ejecución, coordinación y patrocinio. Escribe «no establecido» donde corresponda; no rellenes los huecos con sospechas.
- A2. Explica en 80–120 palabras el mecanismo que pretende utilizar Faro. Incluye por qué la exactitud del horario no elimina el engaño y qué efecto no está demostrado.
- A3. Propón dos pruebas adicionales: una para investigar quién financia a Faro y otra para conocer la recepción. Explica qué respondería cada prueba y qué no bastaría para concluir.

Formato de respuesta

Usa una fila por situación y cuatro columnas: identidad / ejecución / coordinación / patrocinio. Debajo, redacta el mecanismo y las pruebas que faltan. No necesitas buscar información externa: los hechos del ejercicio están en D1–D6 y los conceptos en G2–G7.

Criterio de éxito: puedes sostener conclusiones positivas sobre lo que está probado y reservar otras sin convertir la incertidumbre en una absolución general.

G9 · Actividad 2: corrige una lectura defectuosa

Tiempo orientativo: 10–12 minutos. Los resúmenes S1–S3 son paráfrasis de las fuentes, no citas textuales. El texto I que sigue es ficticio y contiene errores deliberados.

Datos para trabajar

- S1. Unheard Voice describe personas ficticias y coordinación. Las plataformas informaron de países de origen; el estudio no atribuyó las campañas encubiertas a una organización concreta. [F2, pp. 2–3/PDF 3–4]
- S2. Para el conjunto de Asia Central, los autores consideraron probable una adquisición inauténtica de seguidores, posiblemente comprados, a partir de su evolución. [F2, p. 12/PDF 13]
- S3. El ensayo de 2024 adquirió interacción en publicaciones no políticas y comunicó que el 86 % de la interacción artificial permanecía cuatro semanas después. No midió actitudes. [F3, pp. 7, 10]

I · Interpretación que debes revisar

«Los perfiles estudiados eran todos bots y pertenecían sin duda a un gobierno identificado. La compra de sus seguidores quedó probada. El experimento posterior confirma que el 86 % de los usuarios aceptó sus mensajes. Como algunas noticias difundidas eran ciertas, el comportamiento de la red también era auténtico».

Tareas

- B1. Identifica cinco errores. Para cada uno, escribe una corrección y el dato o distinción que la justifica.
- B2. Redacta una conclusión de 90–130 palabras que conserve lo que sí se sabe sobre las conductas y explicite lo que no se sabe. No mezcles las poblaciones de los dos estudios.
- B3. Explica por qué resultados de interacción diferentes en ambos estudios no constituyen por sí mismos una contradicción.

Preguntas para revisar tu respuesta

¿Has confundido activos con personas? ¿Has convertido el país señalado por una plataforma en identidad de un patrocinador? ¿Has tratado una hipótesis como una factura comprobada? ¿Has usado una proporción de actividad como proporción de usuarios convencidos? ¿Has confundido la verdad del contenido con la autenticidad de la participación?

Antes de pasar a G10, subraya en tu respuesta una conclusión respaldada y una limitación. Si todo son afirmaciones absolutas o todo son dudas, revisa la relación entre cada conclusión y su evidencia.

G10 · Soluciones razonadas

Actividad 1

Asociación: miembros identificados, coordinación pública acreditada; no se establece automatización ni patrocinio externo. Oficina: identidad institucional y ejecución automática declarada; eso no prueba inautenticidad. Agencia: identidades ficticias, control común, intervención humana y coordinación encubierta acreditados; el cliente o financiador final no está identificado. Bloguera: seudónimo; D6 no acredita suplantación, automatización ni pertenencia a la red. No saber quién es no autoriza a incorporarla a Faro.

Faro pretende presentar una decisión de empleados como respaldo independiente de vecinos. Así intenta que el apoyo parezca más amplio y espontáneo. La verdad del horario no corrige la biografía ficticia ni la falsa independencia. D4 acredita el propósito expresado y la ejecución humana; D5 no demuestra que vecinos reales recibieran o creyeran los mensajes. Es posible clasificar la conducta sin afirmar persuasión.

Para investigar financiación servirían contratos o pagos autenticados que vincularan a Faro con un cliente y esa campaña. La afinidad temática no los sustituye. Para conocer recepción, podrían examinarse registros de exposición y entrevistar o encuestar a destinatarios identificados. Recordar o ver un mensaje no bastaría para atribuirle un cambio de conducta; esa cuestión exige otro diseño y controlar explicaciones alternativas.

Actividad 2

1. «Todos bots»: personas ficticias y coordinación no prueban automatización de todos los perfiles.
2. «Gobierno identificado»: países señalados por plataformas no equivalen a una organización responsable demostrada; el informe mantiene esa limitación.
3. «Compra probada»: los autores formulan una interpretación, no aportan la transacción como hecho establecido.
4. «86 % de usuarios»: la unidad es interacción artificial persistente en otro estudio; no usuarios ni aceptación.
5. «Noticias ciertas → red auténtica»: veracidad del contenido y representación de identidad e independencia son dimensiones diferentes.

Conclusión posible: «Unheard Voice documentó personas ficticias y conductas coordinadas en campañas encubiertas. Los datos y la investigación abierta sostienen esas observaciones, pero no resuelven plenamente la atribución organizativa. La posible compra de seguidores conserva carácter interpretativo. Por separado, el ensayo de 2024 mostró persistencia de interacción comercial inauténtica en condiciones acotadas. Ese resultado no mide aceptación de mensajes ni puede trasladarse al público de las campañas anteriores. La veracidad de algunas publicaciones tampoco descarta un engaño sobre quién las difunde. Ambos trabajos aportan evidencia sobre organización o actividad, con límites distintos respecto a atribución, alcance y efectos».

No hay contradicción necesaria: una campaña puede recibir poca respuesta espontánea y otro ensayo obtener interacción comprada. Se estudian objetos y condiciones diferentes. Si tu respuesta distingue estos niveles y cita los datos pertinentes, has cumplido el objetivo.

G11 · Fuentes para consultar y síntesis

Las cuatro primeras fuentes pertenecen a la bibliografía proporcionada para el curso; F5 es una consulta adicional. La paginación «PDF» cuenta desde la primera hoja del archivo consultado, incluidas cubiertas. Los títulos permiten localizar la obra aunque cambie su enlace.

[F1] Bradshaw, Samantha; Bailey, Hannah; Howard, Philip N. (2021). Industrialized Disinformation: 2020 Global Inventory of Organised Social Media Manipulation. Oxford Internet Institute, Working Paper 2021.1. Edición proporcionada CyberTroop-Report-2020-v.2.pdf: método pp. 6–7/PDF 8–9; actores pp. 8–10/PDF 10–12; cuentas p. 11/PDF 13; técnicas p. 15/PDF 17. El año del inventario es 2020; el de publicación, 2021.

Oxford Internet Institute · proyecto DemTech

[F2] Graphika y Stanford Internet Observatory (24 agosto 2022). Unheard Voice: Evaluating five years of pro-Western covert influence operations. Resumen y método pp. 2–4/PDF 3–5; Asia Central pp. 11–15/PDF 12–16; figura 10 p. 13/PDF 14.

Graphika · Unheard Voice

[F3] Bergmanis-Korāts, Gundars; Haiduchyk, Tetiana (noviembre 2024). Social Media Manipulation for Sale: 2024 Experiment on Platform Capabilities to Detect and Counter Inauthentic Social Media Engagement. NATO StratCom COE. Método pp. 6–7; resultados pp. 9–10; discusión pp. 25–26. Páginas impresas y PDF coinciden en estos pasajes. ISBN 978-9934-619-07-6.

NATO StratCom COE · editor del informe

[F4] Helmus, Todd C., et al. (2018). Russian Social Media Influence: Understanding Russian Propaganda in Eastern Europe. RAND, RR-2237-OSD. Apartado «Trolls and Bots», pp. 22–24/PDF 40–42. Síntesis contextual de trabajos anteriores, no prueba directa de cada atribución que reseña.

RAND · RR2237

[F5] Gleicher, Nathaniel (16 mayo 2019). Removing Coordinated Inauthentic Behavior From Israel. Facebook Newsroom, sin paginación. Fuente primaria sobre la retirada y la atribución comunicadas por la plataforma.

Facebook · comunicación de la retirada de 2019

Antes del examen

Explica con tus palabras una diferencia entre anonimato y suplantación, otra entre automatización y engaño, y otra entre actividad y efecto. Después identifica qué evidencia prueba la gestión de una red y qué evidencia necesitarías para atribuirla a un patrocinador. Si no puedes justificarlo, vuelve a G2–G7 y contrasta G10.