

LUMBRE



Deepfakes y discurso público: cómo viaja un bulo político, qué defensa corresponde a cada ataque y qué hacer en la primera hora

Materia: Poder público y gobernanza

03/10/2026

Antes de empezar

Objeto 3 de 4, etapa Discurso público, deepfakes y la defensa de la credibilidad. Por qué el discurso se juzga en once segundos, cómo nace y se instala un bulo en setenta y dos horas, los cuatro ataques de información y su defensa (peritaje, versión larga, fuente a la vista, denuncia de patrones), el protocolo de la primera hora y la diferencia entre una crisis negada y una contenida. Cinco videos y cuatro diagramas propios.

QUÉ APRENDERÁS

- Explicar la asimetría entre fabricar y verificar material falso, describir las vueltas del ciclo de un bulo político y su canal de circulación, clasificar cuatro tipos de ataque de información con su defensa correspondiente, aplicar un protocolo de primera hora con archivo y acciones concretas, y escribir el protocolo propio con el archivo mínimo de evidencia de un área.

Discurso público, deepfakes y la defensa de la credibilidad

Este objeto trata de lo que le pasa a la palabra pública cuando fabricar se vuelve barato y probar se vuelve lento. Sirve para un cargo electo que ya vio un audio falso suyo circulando, para una vocería que necesita un protocolo escrito antes de la próxima semana, para un periodista que cubre política local y para cualquier persona que reciba cadenas en un grupo familiar.

Un político siempre fue juzgado por dos cosas distintas: por lo que dijo y por lo que la gente creyó que dijo. La máquina no inventó esa distancia, pero la volvió industrial. Hoy circulan audios que nunca se grabaron, videos que nunca se filmaron y capturas que quizá nunca existieron, y el costo de probar que algo es falso lo paga quien ya tenía poco tiempo y poco personal.

La única defensa real contra eso no es una herramienta: es un historial. La credibilidad es un saldo y no un reflejo, se acumula en discusiones pequeñas y se gasta en una grande, y determina cuántas horas te concede la gente antes de decidir con qué versión se queda. Un cargo público con saldo alto sobrevive a un deepfake; uno con saldo bajo no necesita que lo falsifiquen.

Ficha del objeto · Lumbre

Curso: Políticos y vida pública ante la IA (2026-2034) · Ruta: Ruta del cargo público con la máquina en medio · Etapa: Discurso público, deepfakes y la defensa de la credibilidad · Tema 3 de 4 · Dificultad: intermedia · Duración sugerida: tres horas con videos, cuatro diagramas propios y reactivos con evidencia · Edad referencial: desde 18 años · Prerrequisito: objetos 1 y 2: la campaña y el gobierno con modelos · Autoría: Roadwise Consulting.

1. Qué cambió en el discurso público

Lo primero que cambia es la unidad del mensaje: el discurso ya no se escucha una vez: se escucha en veinte recortes distintos, y cada recorte puede ser sincero en sus palabras y falso en su sentido. Un discurso de cuarenta minutos se resume en once segundos, y esos once segundos son los que la gente comenta, repite y cree.

La asimetría entre acusación y defensa es el corazón técnico del problema: fabricar un audio cuesta minutos, verificarlo cuesta días, y el ciclo de noticias se mueve en horas. Por eso la respuesta útil casi nunca es «eso es falso» dicha rápido: es dar material más completo que el que circula, porque a un fragmento solo lo vence un original.

Seis unidades del discurso público y lo que le hace la circulación fragmentada: la cuarta columna es el daño real

Unidad del mensaje	Cómo circulaba antes	Cómo circula ahora	Qué se pierde
El discurso completo	Una transmisión y una nota	Once recortes y un titular	El contexto que sostenía la frase
La rueda de prensa	Al día siguiente en el medio	En vivo y en fragmentos	El control del orden del día
El audio con un gremio	Reunión y minuta	Doce audios de dos minutos	La precisión de lo acordado
La cifra pública	El boletín con fuente	El número sin origen	La trazabilidad del dato
La respuesta del rival	Al día hábil siguiente	En la misma noche	El tiempo de pensar bien
La corrección	Una nota al pie	Nadie la ve	El cierre del ciclo

1. Una funcionaria lee un informe de resultados en una sesión de cuarenta minutos y al día siguiente el tema público es una frase de ocho segundos. ¿Qué decisión de comunicación explica mejor ese resultado?

- ☐ Que no preparó suficientemente la lectura del informe
- ☐ Que no entregó la versión completa y una cifra con fuente antes de que otro eligiera el recorte
- ☐ Que los medios locales no entienden de política pública
- ☐ Que debió publicar el informe solo por sus canales

La velocidad no es un defecto del periodismo ni de las redes: es la condición en la que hoy se decide qué versión queda.

2. La asimetría central del debate público actual es que fabricar un material falso toma minutos y desmentirlo con evidencia toma días.

- ☐ Verdadero
- ☐ Falso

VIDEO

Comunicación política y redes sociales en la era de la posverdad

Una conferencia larga sobre cómo se reorganizó la conversación política con las plataformas. Útil para este objeto porque describe el antes: buena parte de lo que hoy se atribuye a los modelos ya estaba en marcha con granjas humanas, y la máquina cambió el precio, no la lógica.



Escanea para verlo en Lumbre

https://www.youtube.com/watch?v=7enCob_PMFm

2. Cómo viaja un bulo y por qué el desmentido llega tarde

Un bulo político tiene una vida reconocible de cuatro o cinco vueltas en setenta y dos horas. Nace con forma de captura de chat o de transferencia, porque una imagen no necesita peritaje para convencer; viaja por canales cerrados donde no hay moderación ni rastro; salta a lo público cuando una cuenta grande lo recoge como «lo que circula»; y ahí se disuelve o se instala según cómo se responda.

Vida útil de un bulo en una crisis política

Cuatro vueltas que un contenido falso da en setenta y dos ho

Lumbre · Política

BULO



La quinta vuelta es la que sorprende siempre: el bulo no se murió, estaba esperando una elección.

La cuarta vuelta es la única que controla una autoridad: responder con datos, rápido y por escrito. Las otras tres se ganan con archivo, es decir, con dejar rastro buscable.

La parte que más equipos entienden tarde es el canal. El desmentido se publica en la cuenta oficial y el bulo vive en el grupo vecinal, en el chat familiar y en la columna de un medio chico que nadie lee en la sede de gobierno. No es un problema de alcance: es un problema de trayectoria, porque la desmentida nunca entra al mismo tiempo ni por la misma puerta que el rumor.

Contra una captura de pantalla no hay debate posible: una captura no se desmiente: se desplaza con algo más creíble y más completo. De ahí la rutina de archivo que distingue a un equipo profesional de uno reactivo: cada intervención se guarda con sello de tiempo, cada reunión pública tiene grabación larga, y cada cifra que se dice en una rueda de prensa tiene su hoja con fuente en un folder abierto.

Las cinco vueltas del bulo y qué es eficaz en cada una: la última fila es la que casi todos los equipos intentan y casi nunca logra

Fase del bulo	Qué se puede hacer	Qué no funciona	Ventana de tiempo
Nace como captura	Publicar el original con fecha	Negar sin material	Primeras dos horas
Viaja por canales cerrados	Circular la versión completa ahí mismo	Un comunicado oficial para otro público	Primeras doce horas

Fase del bulo	Qué se puede hacer	Qué no funciona	Ventana de tiempo
Lo recoge una cuenta grande	Dar datos y nombres a ese interlocutor	Pedir retirada sin evidencia	Hora 12 a 24
Se instala o se disuelve	Respuesta escrita con acciones	Responder con enojo o tarde	Hora 24 a 72
Vuelve reempaquetado	Tener el desmentido indexado y buscable	Explicar de nuevo desde cero	Cualquier mes siguiente
Se convierte en creencia	Trabajar con quien aún verifica	Perseguir a quien ya decidió	Después de la semana

Una respuesta con datos, a tiempo y por escrito es lo único que corta el ciclo; el resto de las vueltas se ganan con archivo, es decir, con dejar rastro buscable.

ACTIVIDAD
INTERACTIVA
DISPONIBLE EN
LUMBRE

3. ¿Cuántas veces en el último mes compartiste o comentaste una publicación política antes de verificar de dónde venía?



ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE



VIDEO

Deepfakes y elecciones: ¿estamos perdiendo la verdad?

Fundación Friedrich Naumann conversa el caso electoral con enfoque periodístico. Presta atención a un detalle que casi nadie nombra: en los episodios descritos, el daño no lo causó el material sino el tiempo que la autoridad tardó en publicar el original.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=nk2t600gT5I>

VIDEO

Candidatas generadas con IA: cómo identificar un deepfake y las noticias falsas en elecciones

CNN en Español revisa candidaturas sintéticas y verificación en campaña. Sirve para calibrar una expectativa: la detección técnica es la parte fácil de este problema y la credibilidad previa del que denuncia es la difícil.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=rVmE-LOEDhI>

3. Los cuatro ataques y la defensa que corresponde a cada uno

Cuatro tipos de ataque de información se parecen en la indignación que producen y son distintos en la defensa que requieren, y por eso funciona mal responderles con el mismo mensaje. La fabricación pide peritaje, el fuera de contexto pide la grabación entera, la cifra inventada pide la fuente a la vista y la red de cuentas pide denuncia, no explicación.

Cuatro ataques de información y su defensa

Cada tipo de contenido falso necesita una respuesta distinta

Lumbre · Política

ATAQUES

Fabricación total, audio o video sintético SINTETICO

Se defiende con peritaje y con archivo: grabaciones propias fechadas valen más que una negativa en redes.



Contenido real sacado de contexto RECONTEXT

El más difícil de desmentir porque el material existe: se responde con la grabación completa y con la fecha.



Dato inventado con cifras plausibles CIFRA

Se defiende con la fuente original a la vista. Un número sin año ni organismo pelea contra un número con sello.



Campaña coordinada desde cuentas falsas RED

No es un contenido, es un envío: se documenta el patrón y se denuncia ante la autoridad electoral y la plataforma.



El segundo es el que más daño hace: no se puede desmentir algo que sí ocurrió.

El error más común es responder a los cuatro con el mismo mensaje: la fabricación pide peritaje, el fuera de contexto pide la grabación entera y la red de cuentas pide denuncia, no explicación.

El segundo es el que más daño hace y el que menos se previene: no se puede desmentir algo que sí ocurrió. Un fragmento sacado de contexto usa material verdadero, así que toda la defensa consiste en restituir el entorno, y el entorno pesa menos en una conversación de dos minutos que el recorte que confirma lo que alguien ya sospechaba.

El tercero es el más subestimado, porque una cifra inventada con aspecto plausible circula mejor que un problema real mal explicado. Un número sin año ni organismo pelea contra un número con sello y gana el segundo, salvo que alguien publique la fuente original con la misma rapidez con que se publicó el ataque.

5. Clasifica estos ocho materiales según el tipo de ataque que describe, porque la defensa cambia por completo.

Elementos:

- Un audio sintético de una frase que nunca se dijo
- Un video generado de una reunión inexistente
- Once segundos de una intervención de cuarenta
- Una frase cierta dicha a un gremio distinto
- Un porcentaje sin año ni organismo
- Un gráfico con datos reales y eje manipulado
- Cientas cuentas nuevas repitiendo el mismo texto
- Comentarios generados para simular indignación

Fabricación total	. . .
Real fuera de contexto	. . .
Cifra inventada	. . .
Red de cuentas	. . .

6. A un alcalde le circula un video recortado de una sesión real donde supuestamente admite un favoritismo. ¿Cuál es la respuesta con más probabilidad de funcionar en las primeras horas?

- ☐ Publicar un comunicado negando el contenido del video
- ☐ Difundir la grabación completa de la sesión con marca de tiempo
- ☐ Pedir a las plataformas que retiren el material de inmediato
- ☐ Convocar una rueda de prensa para explicar el contexto

La red de cuentas no busca convencer a nadie: fabrica la apariencia de que muchos ya están de acuerdo, y lo único que la encarece es documentar el patrón, denunciarlo y decir en público que se denunció.

7. Frente a una campaña coordinada desde cuentas falsas, la respuesta eficaz es explicativa y no probatoria: hay que demostrar el patrón y denunciarlo.

- ☐ Verdadero
- ☐ Falso

VIDEO

¿Cómo detectar cuentas falsas y manipulación de bots en redes?

Una guía breve y aplicable sobre la cuarta categoría: señales de cuenta, coincidencias de texto y horarios de publicación. Sirve como lista de chequeo para documentar un ataque coordinado en lugar de solamente quejarse de él.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=jBs0ParEMAQ>

4. La primera hora: protocolo de respuesta

El protocolo de la primera hora tiene un orden contraintuitivo y hay que memorizarlo entero: el primer instinto es el correcto: archivar antes de hablar. Captura, enlace, hora y primer punto de difusión. Quien contesta enojado y sin prueba pierde las dos cosas, el hecho y la versión, y encima deja constancia de que no había nada guardado.

Qué hacer la primera hora cuando aparece un video falso tuyo

Seis pasos de un protocolo que conviene tener escrito antes

Lumbre · Política

PROTOCOLO



Un protocolo que se aprende durante la crisis ya es parte de la crisis.

El orden importa y el primero no es responder: es archivar. Quien contesta enojado y sin prueba pierde las dos cosas, el hecho y la versión.

El segundo principio es anunciar acciones y no adjetivos. Denuncia, peritaje, solicitud de retirada y nombre del responsable: lo abstracto se discute y lo concreto se cree. Un «vamos a demandar» dicho con la denuncia ya presentada cambia el tono de la semana; un «es una mentira más» lo sube.

ACTIVIDAD INTERACTIVA
DISPONIBLE EN LUMBRE

8. Protocolo de la primera hora, con responsables



Para dimensionar el ciclo, el objeto toma el ejemplo de un audio falso atribuido a una regidora: a las 2 horas tenía 180 reenvíos en tres grupos, a las 12 horas lo recogió una cuenta con 40 000 seguidores y a las 26 horas ya era el tema de la sesión de cabildo.

9. En el ejemplo del audio falso de la regidora: ¿cuántas horas tardó el material en pasar de los grupos cerrados a una cuenta grande? Resta las 2 horas iniciales de las 12 horas del salto. (número)

Respuesta: (horas)

10. Ordena los cuatro primeros movimientos de la respuesta a un material falso.

Archivar la evidencia con hora y enlace _____

Clasificar el tipo de ataque _____

Publicar el original o declarar que no existe _____

Anunciar la acción concreta con fecha _____

11. Una dependencia detecta un video sintético del titular a las 21:00 y decide no responder hasta terminar el peritaje oficial, para «no equivocarse». Antes del desenlace: ¿qué es lo más probable que ocurra en las siguientes 48 horas?

- ☐ Que el silencio se interprete como prudencia y el tema baje solo
- ☐ Que el vacío de respuesta se llene con la versión del atacante
- ☐ Que las plataformas retiren el material por no tener réplica
- ☐ Que los medios esperen el peritaje antes de publicar

Reflexiona: ¿Dónde has visto una respuesta correcta llegar demasiado tarde?

Regla de las setenta y dos horas

Toda respuesta pública a un material sospechoso debe cerrarse antes de las setenta y dos horas con un resumen escrito: qué circuló, qué se comprobó, qué se hizo y qué no se sabe todavía. Un ciclo sin cierre no se cierra: se reabre en la siguiente elección con la misma captura.

Con los medios locales la lógica es más simple de lo que parece. Un director que recibe el material completo, con fecha y con la versión larga, tiende a usarlo; un director que recibe una exigencia de rectificación con amenazas tiende a publicar la disputa y no el hecho. La defensa de la credibilidad se gana en buena parte fuera de la cuenta oficial.

El patrón de la negación también es reconocible: se tarda cinco días, se llaman rumores a los hechos, se ataca a quien pregunta para no responder sobre el asunto, se borra la publicación que ya no se puede defender y se termina explicando por qué desapareció la primera explicación. Quien niega todo termina disculpando su propia explicación anterior, y ahí se gasta el saldo del que habla la primera sección.

La diferencia entre una crisis negada y una crisis contenida cabe en cuatro movimientos y no depende de tener la razón. Responder el mismo día con lo que se sabe y con lo que no se sabe, separar el hecho de la interpretación y dejarlo escrito, anunciar una revisión con plazo y con nombre, y asumir el costo de una corrección cuando el dato estaba mal.

5. Crisis negada, crisis contenida y el ejercicio propio

La crisis negada y la crisis contenida

El mismo escándalo digital, dos maneras de manejarlo

Lumbre · Política

DOS CRISIS

X NIEGA

- Tarda cinco días en responder y los llama «rumores».
- Ataca al que pregunta para no responder sobre el hecho.
- Borra la publicación que ya no se puede defender.
- Termina explicando por qué desapareció la primera explicación.



✓ CONTIENE

- Responde el mismo día con lo que sabe y con lo que no sabe.
- Separa el hecho de la interpretación y lo deja escrito.
- Anuncia una revisión con plazo, con nombre y con resultado.
- Gana el derecho a que le crean en la siguiente discusión.



Borrar una publicación deja dos problemas: el original y la desaparición.

La credibilidad no se gasta en el hecho que se reveló: se gasta en el tiempo que tardó la respuesta y en la publicación que se borró para evitarla.

La diferencia entre los dos lados del diagrama no es la gravedad del hecho: es la disposición a pagar el costo de decir la verdad el mismo día. Un equipo que admite un dato mal dicho el martes conserva el beneficio de la duda el martes de la semana siguiente.

El ejercicio de este objeto es escribir dos cosas: el protocolo de la primera hora con nombres y teléfonos, y el archivo mínimo de evidencia que tu área va a sostener durante seis meses. Escribir el protocolo cuando todavía no hay crisis es literalmente la única ventaja que queda, porque durante la crisis nadie aprende: ejecuta.

El archivo mínimo que un área debería sostener seis meses cabe en cuatro carpetas: intervenciones completas con fecha, cifras publicadas con su hoja de fuente, acuerdos con minutas de una página y respuestas a solicitudes de información con su folio.

Cómo cambia la defensa de la credibilidad cuando se prepara

1. Semana 1 **El protocolo escrito** Nombres, teléfonos y una regla: nadie responde sin la clasificación hecha.
2. Semana 2 **El archivo en orden** Cuatro carpetas, accesos definidos y una rutina de respaldo.

3. Mes 2 **El simulacro** Se prueba el protocolo con un caso fingido y se mide cuánto tarda la primera respuesta.
4. Mes 4 **La relación con los medios** Se entrega material completo a quien lo usa, sin exigir cobertura.
5. Mes 6 **El saldo** La primera acusación seria se responde en horas y el tema no llega a la sesión siguiente.

VIDEO

Cómo verificar la desinformación en ciencia y salud

Fundación Gabo enseña la rutina de verificación con un material que no es político y por eso sirve más: metodología de fuente, fecha e interlocutor. Aplícala a la última cadena que recibiste sobre un problema local y anota en cuál de los tres pasos se cayó la conversación.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=NwmUf1aWx7s>

ACTIVIDAD
INTERACTIVA
DISPONIBLE EN
LUMBRE

12. «Ante un deepfake, la mejor estrategia es ignorarlo: responderle le da publicidad.»



13. Contra una captura de pantalla no se discute: se la
_____ . Y a un fragmento solo lo vence un
_____ .

14. Une cada situación con la defensa que le corresponde según el objeto.

14.	Audio sintético inexistente	-----
14.	Frase real recortada	-----
14.	Cifra sin fuente	-----
14.	Cuentas repetidas	-----
14.	Bulo en grupo cerrado	-----
14.	Corrección necesaria	-----

Opciones: Peritaje y grabaciones propias fechadas · La versión larga con sello de tiempo · Publicar la hoja con organismo y año · Documentar el patrón y denunciar · Circular el material ahí mismo · Hacerla visible el mismo día

ACTIVIDAD
INTERACTIVA
DISPONIBLE EN
LUMBRE

15. Sopa de letras de la credibilidad

Ocho palabras del objeto, todas
en horizontal, en mayúsculas y
sin acentos.



ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

16. Tarjetas del objeto



ACTIVIDAD INTERACTIVA
DISPONIBLE EN LUMBRE

17. Memoria del objeto

Empareja cada ataque con el
documento que lo desarma.



ACTIVIDAD INTERACTIVA DISPONIBLE EN
LUMBRE

18. Mi protocolo de la primera hora



Del lado ciudadano la destreza que vale no es detectar deepfakes: es reconocer el propio impulso de compartir. La evidencia disponible es consistente en un punto: quien ya está de acuerdo comparte sin verificar, y la verificación no cambia convicciones,

cambia tiempos. Por eso la única intervención pública con efecto es la que se hace antes de que la pieza llegue a su punto de máxima circulación.

19. Escribe la primera línea del comunicado que tendrías que publicar esta noche si te atribuyeran un audio falso hoy, con tu nombre y tu cargo. Después anota qué evidencia tienes realmente a mano para sostenerla y qué te falta. La diferencia entre las dos cosas es tu trabajo de las próximas dos semanas.

Si la segunda lista está vacía, el problema no es el atacante futuro: es la rutina que todavía no existe.

La defensa de la credibilidad no se fabrica en la crisis: se fabrica en los meses anteriores, acumulando correcciones visibles y guardando el material que prueba lo que uno dijo.

— Idea central del objeto 3

La consecuencia de fondo es política y no comunicacional. Si prueba resulta demasiado caro y negar resulta barato, el incentivo estructural empuja a los cargos públicos a desmentir todo y a no arriesgar nunca una frase verificable. Una democracia no se cae por los deepfakes: se erosiona cuando dejar de decir cosas comprobables se vuelve la estrategia racional.

Un político ante la máquina tiene dos activos que ningún generador puede fabricar por él: un historial de correcciones visibles y un archivo de lo que dijo. El primero hace creíble su negación; el segundo la hace demostrable. Todo lo demás, en la primera hora, es reacción.

Trazabilidad curricular del objeto

Dónde vive este objeto dentro de la ruta del curso

Campo	Valor
Categoría	Poder público y gobernanza
Curso	Políticos y vida pública ante la IA (2026-2034)
Ruta	Ruta del cargo público con la máquina en medio
Etapas	Discurso público, deepfakes y la defensa de la credibilidad

Campo	Valor
Área de conocimiento	Poder público y gobernanza · discurso, verificación y reputación
Tema	Discurso público, deepfakes y la defensa de la credibilidad
Objeto en la ruta	3 de 4
Dificultad	Intermedia
Modalidad	Autoestudio con video, diagrama propio y reactivos con evidencia
Prerrequisito	objetos 1 y 2 de la ruta
Perfil de referencia	Cargo electo y su equipo de comunicación, vocería, periodista local, comunicador social y ciudadano que recibe cadenas por WhatsApp
Autoría	Roadwise Consulting

El objeto 4 cierra el curso mirando hacia adelante: qué le pasa al oficio político y a las reglas del juego entre 2026 y 2034, y qué conviene tener listo antes de que el cambio sea urgente.

— Bajada al objeto 4

Referencia curricular

Materia:	Poder público y gobernanza
-----------------	----------------------------

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://lumbre.lat/object/pol-03>

Material educativo generado desde Lumbre · lumbre.lat