

Estudio sobre La Gobernanza Algorítmica y la Influencia de Intereses Económicos y Políticos en la Era de la Inteligencia Artificial Generativa

Autor: Álvaro Galán Alonso

Fecha: 22/02/2026

Introducción al Paradigma de la Información Sintética y la “Economía de la Atención”

La convergencia histórica entre la economía de la atención y la democratización de la Inteligencia Artificial (IA) generativa ha provocado una transformación tectónica en el ecosistema digital, mediático y sociopolítico. En un periodo caracterizado por la transición desde una era de escasez de información hacia un paradigma de sobreabundancia hipertrófica, la capacidad cognitiva y la atención humana se han erigido como los recursos más valiosos, disputados y escasos del entramado tecnológico contemporáneo.¹ La adopción masiva de modelos fundacionales y herramientas de generación de texto, imagen, audio y vídeo ha inundado las plataformas de comunicación masiva con un volumen sin precedentes de medios sintéticos. Esta hiperabundancia plantea un desafío de dimensiones existenciales para los grandes conglomerados tecnológicos internacionales: cómo preservar la confianza del usuario y la integridad de sus modelos de negocio, los cuales están estrictamente basados en la monetización del compromiso (engagement), sin sofocar la innovación tecnológica ni alienar a los creadores de contenido.

Ante este complejo panorama, empresas dominantes como Alphabet, Meta, ByteDance (TikTok), LinkedIn y X han implementado profundos y divergentes ajustes algorítmicos a lo largo de los años 2024 y 2025. Estas modificaciones arquitectónicas en sus sistemas de recomendación y moderación no responden única y exclusivamente a un imperativo ético o moral de proteger a la ciudadanía de la desinformación estructurada. Fundamentalmente, obedecen a una necesidad estructural y corporativa de defender sus ecosistemas contra el contenido automatizado de baja calidad que amenaza con degradar la experiencia del usuario final, lo que resultaría en un éxodo de la atención y, consecuentemente, en una merma de los ingresos publicitarios.²

Simultáneamente a esta reconfiguración técnica, la intermediación algorítmica de la información ha alterado de forma irreversible el derecho fundamental a la libre información. La selección, jerarquización, distribución y amplificación de noticias y contenidos de interés público ya no obedece estrictamente a criterios de relevancia periodística, rigor empírico o trascendencia cívica. Por el contrario, la dieta informativa de la sociedad está intrínsecamente condicionada por métricas de interacción, intereses corporativos de retención de usuarios, estrategias de financiarización mediática y, en última instancia, presiones políticas y económicas gubernamentales.⁴

El presente documento despliega una investigación exhaustiva y multidimensional que desgrana, en primer lugar, la respuesta técnica, algorítmica y política de las principales

plataformas frente al contenido generado por IA. Posteriormente, profundiza de manera crítica en las dinámicas subyacentes que condicionan la información que la ciudadanía consume diariamente, demostrando cómo los modelos de rentabilidad publicitaria, los sesgos algorítmicos inherentes a las redes neuronales y las estructuras de propiedad mediática —con un enfoque empírico y analítico centrado en el caso de España— subordinan el derecho a la información a intereses sistémicos ajenos al interés público.

Ajustes Algorítmicos y Políticas Defensivas de las Grandes Plataformas

La respuesta arquitectónica de los gigantes tecnológicos frente a la irrupción de la IA generativa ha sido notablemente heterogénea, oscilando en un espectro que abarca desde la asimilación recompensada hasta la desregulación total, pasando por la restricción defensiva y el etiquetado criptográfico. La gobernanza de estos ecosistemas revela un esfuerzo titánico por distinguir el uso legítimo de la automatización para la mejora de la productividad, del abuso manipulativo diseñado para explotar las vulnerabilidades de los motores de búsqueda y los feeds de recomendación.

Alphabet y la Filosofía de la Calidad: SpamBrain y el E-E-A-T

Alphabet, a través de su motor de búsqueda omnipotente Google, ha adoptado una postura sumamente pragmática y orientada a los resultados frente al contenido sintético. En lugar de promulgar una prohibición categórica o un veto algorítmico contra el material generado por IA, la compañía ha reiterado de manera sistemática que sus algoritmos de clasificación están diseñados para recompensar la calidad intrínseca del contenido, independientemente del método tecnológico empleado en su producción.² Esta filosofía fundacional se articula a través del paradigma evaluativo conocido como E-E-A-T (Experiencia, Conocimiento, Autoridad y Confiabilidad, por sus siglas en inglés). La premisa algorítmica de Google sostiene que la automatización puede, de hecho, utilizarse de manera productiva para generar contenido útil y valioso para el usuario final, asimilando este fenómeno contemporáneo a los temores, en gran medida infundados, que surgieron hace una década frente al contenido humano producido en masa mediante granjas de clics.²

No obstante, esta aparente permisividad técnica está estrictamente delimitada por barreras de seguridad algorítmica. Alphabet considera una violación flagrante de sus políticas de spam el uso de IA o cualquier otra forma de automatización con el propósito principal y deliberado de manipular los rankings de búsqueda.² Para hacer cumplir esta directriz a escala global, la corporación ha desplegado actualizaciones algorítmicas masivas durante 2024 y 2025 impulsadas por "SpamBrain", un sistema avanzado basado en aprendizaje automático profundo diseñado específicamente para neutralizar abusos sistémicos.⁶

SpamBrain no se limita a la mera detección de marcas de agua sintácticas propias de las redes neuronales en la redacción, sino que ejecuta un análisis holístico de patrones de comportamiento. El sistema está calibrado para detectar y penalizar tácticas

manipulativas que han proliferado con la democratización de los modelos de lenguaje masivo (LLM), focalizándose en cuatro áreas críticas de abuso: el secuestro y abuso de dominios caducados para aprovechar su autoridad histórica, la manipulación de la reputación de sitios web legítimos mediante la inyección de contenido automatizado, la creación de contenido a escala industrial sin supervisión humana, y la publicación de contenido de bajo esfuerzo que carece de valor semántico original.⁶ Las actualizaciones más recientes, particularmente la implementada en agosto de 2025, han integrado modelos de IA capaces de interpretar la intención del usuario con una precisión sin precedentes, desplazando el peso algorítmico desde la coincidencia exacta de palabras clave hacia la evaluación del contexto, y penalizando tácticas obsoletas de manipulación semántica, encubrimiento (cloaking) y esquemas de enlaces artificiales.⁷

Paradójicamente, mientras Google rastrea y penaliza implacablemente el contenido sintético de terceros que carece de valor añadido humano, ha integrado de forma agresiva su propia tecnología de IA generativa en la cúspide de sus resultados de búsqueda mediante las denominadas "AI Overviews" (Resúmenes de IA). Este movimiento estratégico ha alterado drásticamente la arquitectura del tráfico web global, priorizando la retención del usuario dentro del ecosistema cerrado de Alphabet frente a la derivación de tráfico orgánico hacia los creadores y editores originales.⁸

Característica del Ecosistema de Alphabet	Impacto en la Distribución de la Información
Evaluación E-E-A-T	Prioriza la autoridad demostrable; la IA sin revisión humana experta es relegada a posiciones inferiores. ²
Algoritmo SpamBrain	Neutraliza el "scaled content abuse"; los dominios que automatizan publicaciones masivas pierden indexación. ⁶
Actualizaciones Core 2025	Evolución hacia modelos basados en intención (BERT, MUM) que devalúan el relleno de palabras clave y el SEO tradicional. ⁵
AI Overviews	Generación de respuestas directas ("Zero-click"); devaluación drástica del tráfico de referencia hacia editores independientes. ¹⁰

Meta y la Transición hacia la Contextualización y el Etiquetado Modular

El enfoque algorítmico y de políticas públicas de Meta (matriz de Facebook, Instagram y Threads) ha experimentado una evolución metodológica y filosófica significativa, transitando desde la eliminación punitiva hacia la contextualización informativa y la máxima preservación de la libertad de expresión. Hasta mediados del año 2024, las políticas de medios manipulados de la compañía se centraban casi exclusivamente en eliminar videos "deepfakes" en los que una persona aparecía diciendo algo que empíricamente no había dicho. Sin embargo, tras extensas revisiones y consultas directas con el Consejo Asesor de Contenido (Oversight Board) y un amplio abanico de expertos de la industria en más de 34 países, Meta concluyó formalmente que esta política era excesivamente estrecha, obsoleta e ineficaz en un ecosistema digital dominado cada vez más por imágenes hiperrealistas y audios sintéticos.¹²

En respuesta a este diagnóstico, la nueva arquitectura algorítmica de Meta prioriza el contexto, sustituyendo la eliminación sistemática de contenidos por un complejo sistema de etiquetado en capas. Inicialmente, la plataforma implementó la etiqueta general "Made with AI" (Creado con IA). Sin embargo, ante la masiva retroalimentación de fotógrafos y usuarios cotidianos, quienes percibían esta categorización como injusta e inexacta para fotografías legítimas que solo habían recibido ajustes menores de iluminación, enfoque o recortes mediante herramientas de IA integradas en los teléfonos móviles, la compañía migró en julio de 2024 a la etiqueta semánticamente más matizada de "AI info" (Información de IA).¹²

Para la implementación técnica de estas etiquetas, los algoritmos de ingesta de Meta confían en la detección de señales compartidas por la industria (estándares técnicos comunes invisibles incrustados en los archivos), así como en la autodeclaración voluntaria de los propios usuarios.¹² A partir de septiembre de 2024, el algoritmo introdujo una distinción crucial basada en el grado de manipulación detectada: si el contenido ha sido meramente modificado o editado superficialmente por herramientas de IA, la advertencia "AI info" se relega discretamente al menú secundario de la publicación para no interferir con la experiencia visual. Por el contrario, si el sistema detecta que el contenido ha sido generado íntegramente de forma sintética, la etiqueta permanece anclada y visible de forma prominente en la interfaz principal.¹²

Esta recalibración algorítmica defensiva se complementa con un viraje estratégico en la moderación de discursos a gran escala. A principios de 2025, Meta anunció en Estados Unidos el fin progresivo de su programa global de verificadores de datos externos (third-party fact-checking), transicionando hacia un modelo descentralizado de "Notas de la Comunidad" (Community Notes) fuertemente dependiente del escrutinio y consenso de los propios usuarios de la plataforma.¹⁴ La directiva de la corporación ha levantado explícitamente restricciones algorítmicas sobre múltiples temas que forman parte del discurso político dominante, enfocando sus recursos computacionales y de revisión humana exclusivamente en violaciones de alta gravedad y contenido manifiestamente

ilegal, como el abuso infantil o la incitación a la violencia inminente.¹⁴

Esta estrategia bifronte logra dos objetivos corporativos fundamentales: reduce drásticamente los gigantescos costes operativos de moderación que Meta había sostenido en años anteriores y fomenta un mayor volumen de interacción libre en la plataforma. Al delegar la responsabilidad de la veracidad epistémica a la inteligencia colectiva, Meta se blindo legalmente y abraza un modelo que, si bien incrementa la libertad de expresión, históricamente ha mostrado profundas vulnerabilidades frente a campañas coordinadas de desinformación impulsadas por actores estatales o grupos de presión.¹² Adicionalmente, en el ámbito comercial, Meta exige a los anunciantes que revelen obligatoriamente el uso de IA cuando alteran digitalmente anuncios relacionados con temas sociales, elecciones o política, asegurando que la monetización publicitaria no cruce la línea de la manipulación electoral encubierta.¹²

ByteDance (TikTok): Trazabilidad Criptográfica, Alfabetización y Control de Interfaz

En agudo contraste con las plataformas occidentales, TikTok ha adoptado uno de los marcos de trazabilidad algorítmica y divulgación más rigurosos, técnicos y proactivos de toda la industria tecnológica. Reconociendo tempranamente que el contenido sintético hiperrealista posee el potencial de desdibujar peligrosamente la línea entre la realidad y la ficción en un formato de vídeo corto de alto impacto emocional, la plataforma exige de manera normativa y estricta a sus creadores que etiqueten todo el material generado por IA que simule ser realista.¹⁶

A nivel de infraestructura técnica profunda, TikTok ha sido pionera al convertirse en la primera plataforma de intercambio de vídeos en integrar completamente los estándares de la Coalición para la Procedencia y Autenticidad del Contenido (C2PA). Esta integración tecnológica subyacente permite a los algoritmos de procesamiento de la plataforma leer "Credenciales de Contenido" (metadatos criptográficos a prueba de manipulaciones) y aplicar automáticamente etiquetas inamovibles a los vídeos, audios e imágenes generados por IA que provienen de otras plataformas o generadores de terceros.¹⁶ Gracias a esta sinergia intersectorial, TikTok ha logrado identificar y etiquetar proactivamente más de 1.300 millones de vídeos sintéticos hasta la fecha.¹⁷

Además, para asegurar la trazabilidad de la procedencia del contenido creado dentro de su propia aplicación, la red social está aplicando técnicas avanzadas de "marcas de agua invisibles". Este método de esteganografía digital es extremadamente resistente a la manipulación, recortes o compresión de archivos, permitiendo a los algoritmos rastrear el origen de los medios sintéticos producidos con herramientas nativas, como su potente TikTok AI Editor Pro.¹⁷

Lejos de tratar a sus usuarios como meros receptores pasivos, TikTok ha ejecutado un movimiento estratégico que transfiere el poder de curación algorítmica a la heurística del usuario individual. En 2025, la compañía implementó un control de contenido de

vanguardia dentro de su función "Administrar temas" (Manage topics). Esta herramienta empodera a las personas para que ajusten, mediante un control deslizante intuitivo, la frecuencia y la cantidad exacta de contenido generado por IA que desean que el motor de recomendación inyecte en su feed principal "Para Ti".¹⁷ En lugar de adoptar un enfoque binario y paternalista de censura o eliminación, TikTok ha optado por clasificar el contenido sintético como una categoría temática legítima más, análoga a los deportes, la moda o la gastronomía. De este modo, se normaliza la IA como un género de entretenimiento válido, garantizando simultáneamente una transparencia absoluta.

Para complementar estas medidas arquitectónicas, la plataforma reconoce su papel en el condicionamiento cognitivo de las nuevas generaciones y ha invertido un fondo educativo de 2 millones de dólares destinado a fomentar la alfabetización mediática. A través de asociaciones estratégicas con organizaciones como GirlsWhoCode y MediaWise (un programa del Instituto Poynter), TikTok busca educar a su base de usuarios más joven sobre la seguridad en la IA, el funcionamiento de los algoritmos de recomendación y la identificación de sesgos en los medios digitales.¹⁷

LinkedIn y la Búsqueda Ontológica de la Profundidad Profesional

A diferencia de las plataformas centradas en el entretenimiento general y el consumo rápido, LinkedIn (propiedad de Microsoft Corporation) enfrenta un riesgo existencial profundamente asimétrico: que la IA generativa y los modelos de lenguaje masivo (LLM) inunden su ecosistema profesional con publicaciones corporativas estandarizadas, consejos huecos y textos carentes de valor empírico genuino. En respuesta a las quejas generalizadas de sus usuarios sobre la proliferación de contenido irrelevante y tácticas de autopromoción encubiertas, la red profesional ha ejecutado una de las reestructuraciones algorítmicas más agresivas de su historia durante 2024 y 2025.¹⁸

La alteración fundamental radica en el abandono de las métricas de volumen de interacción superficial (likes rápidos y alcance pasivo) en favor de una nueva heurística computacional que la compañía denomina "profundidad de interacción" (engagement depth).¹⁹ El motor de recomendación de LinkedIn ha sido recalibrado para evaluar meticulosamente la correlación entre la publicación y la "experiencia en la materia" (Subject Matter Expertise) del autor.¹⁸ En términos prácticos, si el sistema detecta que un usuario ha utilizado un LLM para redactar consejos de liderazgo empresarial altamente optimizados, pero dicho usuario carece de un historial profesional verificable, años de experiencia o credenciales en esa industria específica dentro de su perfil, el algoritmo restringe severamente el alcance orgánico de la publicación, clasificándola como contenido de bajo valor añadido.¹⁸

Factor Algorítmico en LinkedIn (2025)	Impacto en la Visibilidad del Contenido
--	--

Experiencia en la Materia	Correlación algorítmica entre el historial del perfil y el tema publicado; penaliza a "falsos gurús" apoyados en IA. ¹⁸
Relevancia del Interés	Distribución hiper-segmentada a nichos específicos en lugar de difusión masiva (broadcasting) inespecífica. ¹⁸
Profundidad de Interacción	Priorización absoluta de comentarios extensos y debates sustanciales sobre las reacciones de un solo clic. ¹⁹
Fin de Públicos Similares	Eliminación de "Lookalike Audiences" (Feb 2024), forzando a los anunciantes a construir audiencias basadas en interés orgánico real. ²⁰

El sistema de distribución ha sido reconfigurado topológicamente. Se otorga una prioridad absoluta a la visibilidad entre las conexiones de primer grado y al público profesional altamente segmentado, penalizando algorítmicamente el contenido de "cebo" diseñado puramente para atraer interacciones de baja calidad, así como las actualizaciones personales de estilo de vida publicadas por perfiles ajenos a la red inmediata del usuario.¹⁸ Con la eliminación definitiva de la función publicitaria de "Públicos Similares" (Lookalike Audiences) a principios de 2024, LinkedIn ha clausurado una vía de hiper-optimización, obligando a los creadores, empresas B2B e instituciones a construir su influencia basándose en la entrega de conocimiento auténtico, perspectivas únicas y experiencia vivida, activos que la IA generativa aún no puede emular convincentemente.²⁰

X y la Desregulación Sistémica del Entorno Sintético

La estrategia algorítmica y corporativa de X (anteriormente conocida como Twitter) diverge de manera radical y paradigmática del consenso protector de la industria tecnológica. Bajo su actual modelo de gobernanza, la plataforma ha optado por implementar su propio modelo generativo nativo, denominado Grok, integrándolo profundamente en la estructura y la experiencia de los suscriptores premium (X Premium). A diferencia de los modelos fundacionales desarrollados por OpenAI, Anthropic o Google, que incorporan estrictos guardarraíles éticos, equipos de "red teaming" y filtros de seguridad en su arquitectura base, el ecosistema de X se ha caracterizado por una política deliberada de extrema permisividad técnica que ha maximizado deliberadamente la generación de contenido altamente controvertido, sin

censura previa.²¹

A finales de 2024 y durante gran parte de 2025, las actualizaciones en las capacidades multimodales y de generación de imágenes de Grok permitieron a los usuarios crear representaciones visuales que eludían fácilmente las restricciones tradicionales de la industria. Las instrucciones del sistema de Grok permitían explícitamente contenido para adultos, temáticas oscuras e imaginaria violenta bajo el pretexto de la no censura y la creatividad irrestricta.²² Esto resultó en un aumento exponencial y descontrolado de medios sintéticos en la línea de tiempo de la plataforma. La red se inundó de material protegido por derechos de autor intelectual, desinformación política hiperrealista generada para influir en debates electorales y, de forma críticamente alarmante, la creación masiva y no consensuada de imágenes íntimas simuladas (deepfakes pornográficos) de figuras públicas prominentes, políticas y ciudadanas anónimas.²³

Investigaciones independientes y reportes de organizaciones internacionales como Human Rights Watch documentaron meticulosamente que, en un corto periodo de nueve días tras la activación de ciertas capacidades de edición de imágenes, se generaron aproximadamente 4,4 millones de imágenes sintéticas en la plataforma, de las cuales casi la mitad contenían material explícito o altamente sexualizado, frecuentemente sin el consentimiento de las personas reales utilizadas como modelo o indicación.²⁴

En lugar de implementar filtros algorítmicos robustos, auditorías de seguridad en los datos de entrenamiento o clasificaciones automatizadas de extremo a extremo para purgar el contenido difamatorio, la respuesta técnica y ejecutiva de X se centró en la restricción comercial y el cumplimiento reactivo. En lugar de abordar la falla sistémica y ética del modelo algorítmico subyacente, X optó por limitar el acceso a la herramienta exclusivamente a suscriptores de pago e implementó geobloqueos funcionales en jurisdicciones específicas donde la generación de dicho contenido acarrea responsabilidades penales explícitas.²⁴

Esta desregulación algorítmica intencionada ha desencadenado una severa respuesta institucional. La plataforma enfrenta actualmente investigaciones formales bajo la Ley de Servicios Digitales (DSA) impulsadas por la Comisión Europea, centradas específicamente en el funcionamiento opaco de sus sistemas de recomendación, la propagación de discurso de odio algorítmicamente amplificado y el diseño de su modelo Grok.²¹ De forma paralela, en Estados Unidos, la fiscalía general del Estado de California y una coalición de 35 fiscales generales han iniciado investigaciones sobre la falta de barreras de seguridad de xAI para prevenir la creación de material abusivo y acoso digital.²³

La Eficacia Epistémica de las Etiquetas y la Trampa de las Burbujas Algorítmicas

La inmensa proliferación de ajustes algorítmicos y políticas de moderación plantea una interrogante empírica fundamental para la sociología y las ciencias de la comunicación

en 2025: ¿son realmente eficaces las etiquetas de IA para modular el comportamiento del usuario y frenar la propagación de desinformación? La literatura científica más reciente ofrece conclusiones profundamente contraintuitivas que iluminan la compleja psicología del consumidor de medios digitales en la era del capitalismo de plataformas.

Estudios de Eficacia e Intervención Cognitiva

Investigaciones académicas rigurosas conducidas a lo largo de 2024 y 2025 demuestran que las advertencias visuales y las etiquetas sobre el contenido sintético logran alterar la percepción epistémica a nivel individual, pero fracasan sistemáticamente en alterar el comportamiento de interacción o amplificación social. Un estudio exhaustivo (Berinsky et al., 2025), basado en encuestas pre-registradas a más de 7.500 ciudadanos estadounidenses, evaluó empíricamente dos estrategias distintas de etiquetado en entornos digitales: un enfoque "basado en el proceso" (que simplemente aclara de manera neutral que el contenido fue generado algorítmicamente con IA) y un enfoque "basado en el daño" (que advierte activamente sobre el potencial del contenido para manipular o engañar al espectador).²⁵

Los hallazgos empíricos revelaron que ambos tipos de etiquetas lograron su objetivo primario: disminuyeron de manera estadísticamente significativa la creencia de los participantes en la veracidad fáctica de las afirmaciones y las imágenes presentadas. Sin embargo, en una demostración palpable de la disonancia cognitiva que impera en los entornos hiper-conectados, estas mismas etiquetas tuvieron poco o nulo impacto en la probabilidad declarada por los encuestados de interactuar con la publicación (dar "me gusta", comentar o compartirla con sus redes).²⁵

Este comportamiento disonante subraya un fenómeno psicológico y sociológico crítico provocado por las arquitecturas de las plataformas: la validación identitaria y la señalización tribal a menudo superan con creces a la verificación factual. Los usuarios en redes sociales no comparten información de manera compulsiva únicamente porque creen que es ontológicamente cierta, sino porque el contenido sirve como una poderosa señal de pertenencia a un grupo, afiliación política o indignación moral compartida. Otros estudios experimentales complementarios, como los conducidos por Lu y Yuan (2024) sobre deepfakes de políticos discutiendo el cambio climático, confirman que, si bien las exenciones de responsabilidad claras sobre la IA mejoran sustancialmente el reconocimiento de parodias, la conciencia del engaño no frena completamente el impulso algorítmico y emocional hacia la viralidad impulsada por el sesgo de confirmación.

De las Cámaras de Eco a las "Burbujas Generativas"

El condicionamiento de la información y la restricción del libre pensamiento también se manifiesta de manera endógena y sutil a través de los sesgos inherentes al diseño matemático de los algoritmos de IA. Ningún sistema informático complejo es verdaderamente neutral ni objetivo; invariablemente reflejan las visiones del mundo, las

lagunas y carencias en las inmensas bases de datos de entrenamiento, y los prejuicios cognitivos de los ingenieros que los diseñan. La persistente falta de diversidad demográfica en los grandes conglomerados tecnológicos contribuye irremediabilmente a la cristalización matemática de sesgos sistémicos.

Históricamente, los algoritmos de recomendación de las redes sociales crearon las infames "burbujas de filtro" (filter bubbles) y "cámaras de eco", entornos digitales asfixiantes al exponer a los usuarios única y exclusivamente a perspectivas ideológicamente afines, amplificando el sesgo de confirmación, limitando drásticamente la exposición a la diversidad de puntos de vista y exacerbando la polarización política.²⁶ En 2025, la comunidad científica ha identificado una evolución aún más insidiosa de este fenómeno: el advenimiento de las "burbujas generativas".³³

Las burbujas generativas ocurren y se solidifican cuando las interacciones reiteradas de los usuarios con agentes conversacionales de IA (como ChatGPT, Claude o Gemini) estrechan progresivamente la cosmovisión y el marco conceptual del individuo.³³ Debido a que los grandes modelos de lenguaje están matemáticamente optimizados para predecir y generar la respuesta estadísticamente más probable, coherente y placentera para el usuario (una forma extrema de alineación comercial y satisfacción del cliente), tienden a reafirmar preconcepciones, prejuicios y sesgos previos. El peligro reside en que, a diferencia de un post de redes sociales de dudosa procedencia, la IA presenta estas realidades distorsionadas o información parcial de una manera altamente estructurada, persuasiva, articulada y aparentemente enciclopédica, dotando a la desinformación de una pátina de autoridad inmerecida e infundiendo una falsa sensación de objetividad.³³

La Arquitectura Institucional y el Marco Regulatorio (El Caso Europeo y Español)

Frente a las evidentes y documentadas vulnerabilidades de la autorregulación corporativa, donde los incentivos financieros de las plataformas colisionan directamente con la salud del ecosistema informativo, los legisladores internacionales han respondido erigiendo marcos normativos de obligado cumplimiento.

A nivel supranacional, la histórica Ley de Inteligencia Artificial de la Unión Europea (AI Act) establece en su Artículo 50 obligaciones estrictas y vinculantes de transparencia y trazabilidad.³⁵ Los proveedores de sistemas de IA generativa deben garantizar por diseño que el contenido sintético producido sea detectable mediante formatos legibles por máquina. A su vez, los implementadores (deployers) y las plataformas de distribución masiva tienen el deber fiduciario y legal de revelar de forma inequívoca cuándo se ha utilizado IA para crear contenido sintético realista, advirtiendo claramente a los ciudadanos de que están ante un "deepfake". La normativa solo establece excepciones muy acotadas y minimalistas para salvaguardar la libertad de expresión en obras de naturaleza evidentemente satírica, artística o de ficción manifiesta, buscando un delicado equilibrio entre la protección contra la manipulación y la integridad creativa.³⁵

En el Reino de España, este complejo mandato regulatorio europeo se aterriza y canaliza a través de la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), constituida pioneramente a finales de 2023. La Estrategia de IA 2024 del Estado español posiciona a esta agencia gubernamental como el árbitro central y el actor clave encargado de definir las buenas prácticas sectoriales, auditar los modelos de alto riesgo y garantizar el cumplimiento de las normativas de transparencia algorítmica.³⁹

No obstante, en un esfuerzo por desopacar la gobernanza tecnológica de la propia administración pública, organizaciones independientes de la sociedad civil, agrupadas en la coalición "IA Ciudadana", han presentado informes en 2025 exigiendo la creación de un registro de algoritmos de IA de carácter nacional, público, accesible y obligatorio.⁴¹ Este instrumento cívico busca transparentar detalladamente qué sistemas automatizados se despliegan en la administración (en áreas críticas como sanidad, asistencia social o justicia), con qué propósito específico y con qué efectos sociodemográficos documentados, con el fin de mitigar el alto riesgo de opacidad tecnocrática y cajas negras institucionales que vulneren derechos fundamentales.⁴¹

De forma paralela y complementaria, la Ley General de Comunicación Audiovisual de España ha sido profundamente actualizada para reflejar la nueva realidad del mercado de la atención. El nuevo texto legislativo establece un terreno de juego equilibrado en términos de obligaciones, exigiendo a los nuevos operadores digitales (incluyendo influencers y creadores de contenido nativos de redes sociales) los mismos estándares de transparencia en los flujos informativos, divulgación publicitaria clara y responsabilidades de difusión algorítmica que históricamente se exigían exclusivamente a las cadenas de televisión lineal tradicionales.⁴³

Intereses Corporativos y la Condicionalidad Estructural de la Información

Más allá de la amenaza inmediata que representa la falsificación sintética de la realidad, la cuestión central y más insidiosa del ecosistema digital actual radica en comprender cómo las inmensas infraestructuras algorítmicas de difusión masiva condicionan de forma sistémica y latente la información periodística que recibe la ciudadanía. La promesa utópica original de internet como un ágora horizontal y un espacio de libre flujo informativo ha sido completamente suplantada, subyugada y mercantilizada por sofisticados modelos de negocio corporativos, en los que la atención humana es extraída y comercializada a escala industrial.¹

La Tiranía de la Economía de la Atención y el Criterio de "Shareworthiness"

Los algoritmos de recomendación (news feeds, For You pages) no operan en un vacío ético o neutral; son construcciones matemáticas altamente intencionadas, diseñadas de forma iterativa para maximizar objetivos comerciales corporativos, siendo los más preeminentes el tiempo de retención en pantalla (dwell time) y la expansión del inventario publicitario.⁴ Revisiones sistemáticas e investigaciones empíricas publicadas a lo largo de 2025 demuestran de forma concluyente que los sistemas algorítmicos de las

redes sociales han reconfigurado y erosionado el concepto fundamental y centenario del valor periodístico.

La "noticiabilidad" (newsworthiness) histórica, fundamentada tradicionalmente en factores como la trascendencia social del evento, el rigor empírico, la precisión de los hechos y la utilidad cívica para una ciudadanía informada, ha sido desplazada inexorablemente por la dictadura de la "compartibilidad" (shareworthiness).⁴

Las arquitecturas subyacentes de las plataformas privilegian y amplifican de manera desproporcionada la resonancia emocional extrema, la controversia, el conflicto tribal y la indignación moral, puesto que los datos demuestran empíricamente que son los vectores psicológicos más eficientes para retener la atención humana y provocar interacciones impulsivas. Este condicionamiento algorítmico altera profundamente la autonomía editorial de los medios de comunicación tradicionales y digitales: las redacciones de noticias se enfrentan a lo que la sociología de los medios define como una "agencia limitada" (bounded agency).⁴

Atrapados en esta arquitectura asimétrica, los profesionales de la información se ven forzados a practicar una perniciosa "autocensura anticipatoria".⁴ Es decir, los periodistas y editores evitan sistemáticamente abordar temas de alta complejidad técnica, análisis geopolíticos profundos o reportajes de tono moderado y conciliador, albergando la certeza matemática de que dichos contenidos serán penalizados y sepultados por los motores de recomendación. En su lugar, se prioriza la producción industrializada y rápida de contenido de baja calidad optimizado para la viralidad instantánea. Esta dependencia parasitaria y extrema de las métricas comerciales impuestas por Silicon Valley degrada la legitimidad institucional de la prensa como "perro guardián" de la democracia, precariza las condiciones laborales del sector e intensifica aceleradamente la polarización política de la sociedad.⁴

La Apropiación Extractiva del Tráfico: El Efecto Destructivo de las "AI Overviews"

El conflicto frontal entre el derecho a la libre información, la viabilidad del periodismo independiente y los intereses monopolísticos corporativos se evidencia de manera aguda y letal en la reciente implementación de la IA generativa en las interfaces de los motores de búsqueda, con Alphabet a la vanguardia.

La introducción y expansión de las "AI Overviews" por parte de Google ha alterado radicalmente y puesto en riesgo inminente la frágil economía de los medios de comunicación digitales. Al utilizar LLMs para raspar (scrape), ingerir y sintetizar respuestas definitivas directamente en la página de resultados de búsqueda —un fenómeno que la industria ha bautizado como la consolidación del "Zero-click internet" o internet de cero clics— la plataforma se apropia unilateralmente del inmenso valor intelectual y financiero del contenido periodístico original.⁷ De este modo, Google logra su objetivo comercial supremo: retener al usuario dentro de sus fronteras digitales,

destruyendo simultáneamente el tráfico de referencia orgánico del cual dependen los creadores, editores y medios de comunicación para su viabilidad y supervivencia financiera.¹⁰

Los datos empíricos agregados de la industria tecnológica e informes de medios durante 2024 y 2025 reflejan la magnitud devastadora de esta disrupción arquitectónica, revelando caídas dramáticas y sostenidas en el Click-Through Rate (CTR) de las publicaciones de noticias y artículos informativos:

Fuente del Estudio de Mercado	Métrica Analizada frente a la Intervención de "AI Overviews"	Descenso Estimado del Tráfico / CTR
Ahrefs (Dic. 2025)	Impacto en la Tasa de clics (CTR) en las páginas mejor posicionadas tradicionalmente	- 58%
Seer Interactive	Descenso de clics orgánicos globales para consultas puramente informativas	- 49.4% a - 65.2%
Digital Content Next (DCN)	Caída del tráfico de referencia interanual general en un panel de grandes medios premium	- 10% a - 25% (reportando picos de pérdida del -75% en ciertas categorías)
Pew Research Center	Comportamiento del usuario: proporción que hace clic en enlaces tradicionales con presencia de IA	Desplome drástico del 15% histórico al 8% actual
SearchEngineLand	Disminución del CTR orgánico para consultas informativas desde mediados de 2024	- 61%

Tabla 1: Impacto financiero y analítico de las interfaces de IA generativa integradas en los motores de búsqueda sobre el tráfico de referencia vital para el sostenimiento de los

creadores de contenido y la prensa.¹⁰

La respuesta a esta asfixia estructural no se ha hecho esperar. Grandes conglomerados editoriales, plataformas de contenido educativo como Chegg Inc., y editores de revistas de prestigio mundial como Penske (propietaria de Rolling Stone y Billboard) han interpuesto demandas legales históricas y denuncias antimonopolio. Estas corporaciones argumentan que Alphabet y otras plataformas tecnológicas abusan de su posición de monopolio indiscutible para forzar a los creadores a entregar sus contenidos originales como condición ineludible para mantener la visibilidad en las búsquedas, compitiendo deslealmente en el mercado de provisión de contenido.⁸ Sin embargo, juristas y académicos advierten que estas demandas enfrentan un obstáculo legal complejo: la pretensión de establecer un "derecho al tráfico", argumentando que el clic no es una forma de propiedad intelectual codificada en la jurisprudencia, lo que enmaraña aún más la defensa del periodismo independiente.⁸

Agravando sustancialmente esta crisis de viabilidad, análisis semánticos y técnicos exhaustivos de las citas proporcionadas por estos sistemas generativos demuestran un claro y pernicioso sesgo algorítmico hacia la autorreferencia corporativa. Evaluaciones realizadas sobre más de 36 millones de "AI Overviews" revelan que Google tiende a enlazar de manera abrumadoramente mayoritaria hacia sus propias propiedades comerciales (como YouTube) y hacia foros comunitarios de contenido generado por usuarios (UGC) como Reddit. De hecho, el 38% de todas las citas generadas por IA provienen de un oligopolio de cinco dominios (Wikipedia, YouTube, Reddit, propiedades internas de Google y Amazon), mientras que la plataforma exhibe un comportamiento de enlace autorreferencial del 43%.⁹ Este control oligopólico y endogámico sobre la curación y distribución de la información margina sistemáticamente al periodismo profesional de investigación, subordinando el acceso público a datos rigurosos al imperativo financiero innegociable de mantener al usuario cautivo y monetizable dentro del feudo corporativo de la plataforma.

Intereses Políticos y Estructuras Económicas: El Caso de Estudio de España

Para comprender con total profundidad hasta qué punto la información diaria que la sociedad civil recibe está transversalmente condicionada por intereses distintos al libre flujo de conocimiento democrático, es absolutamente imprescindible analizar la economía política subyacente de los medios de comunicación a nivel nacional. El ecosistema informativo de España en 2025 ofrece un caso de estudio paradigmático y alarmante sobre la intersección crítica entre la hiper-concentración de la propiedad empresarial, la precarización financiera crónica estructural, la disrupción tecnológica externa y la intervención y presión directa del aparato estatal.⁵⁰

Concentración Oligopólica y la Financiarización de las Redacciones

A contrapelo de la ilusión de pluralidad que aparenta internet, el panorama mediático español está férreamente dominado por un puñado de conglomerados fuertemente

consolidados, históricamente blindados y entrelazados con las élites económicas del país: Grupo PRISA, Grupo Vocento, Grupo Planeta y Mediaset España.⁵⁰ Durante las últimas dos décadas, estos poderosos grupos corporativos han implementado estrategias implacables de convergencia mediática, controlando simultáneamente y de manera sinérgica la prensa escrita de tirada nacional, inmensas redes de emisoras de radio, cadenas de televisión generalistas y una constelación de plataformas digitales.⁵² El propósito fundamental de esta integración vertical y horizontal ha sido lograr economías de escala drásticas y reducir los costes de producción periodística, todo ello en un intento desesperado por contrarrestar la hemorragia masiva de ingresos publicitarios absorbidos casi en su totalidad por el duopolio tecnológico global (Google y Meta).⁵²

Esta incesante consolidación empresarial ha venido acompañada de un agudo y transformador proceso de "financiarización" mediática. La entrada masiva de fondos de inversión internacionales de perfil buitre, bancos acreedores nacionales (cuyos préstamos mantienen a flote a las empresas) y diversas instituciones financieras opacas en el accionariado mayoritario de los principales medios de comunicación españoles ha transmutado subrepticamente, pero de manera innegable, las líneas editoriales de las cabeceras históricas.⁵⁰

En este paradigma corporativo, cuando las obligaciones implacables de la deuda financiera y la exigencia de retornos trimestrales a los accionistas institucionales prevalecen jerárquicamente sobre el mandato cívico y democrático del periodismo, las líneas de investigación en las redacciones son invariablemente sofocadas o condicionadas. Los directores de medios se ven abocados a evitar investigaciones profundas o posturas editoriales que entren en conflicto directo con los intereses económicos de los grandes corporativos, los fondos de inversión tenedores de la deuda o los monopolios energéticos e inmobiliarios que fungen como sus principales anunciantes.⁵¹ Las colosales pérdidas estructurales continuadas que sufre el sector —ilustradas dramáticamente por grupos editoriales reportando caídas financieras por valor de decenas de millones de euros y ejecutando expedientes de regulación de empleo (ERE) sistemáticos— multiplican exponencialmente la vulnerabilidad y la debilidad de las redacciones periodísticas ante cualquier factor de coacción o presión externa, disolviendo el blindaje ético tradicional del periodista.⁵³

Intervención Gubernamental, Subsidios Institucionales y el Eje de Subsistencia

Esta asfixiante fragilidad económica de carácter crónico ha convertido a las empresas periodísticas españolas en entidades extremadamente dóciles y peligrosamente dependientes de las inyecciones de capital provenientes del poder político, los gobiernos autonómicos y la administración central. La materialización más evidente de esta relación de sumisión se observa en el controvertido "Plan de Acción por la Democracia" implementado por el gobierno español en el periodo de 2024 y 2025. Este programa estatal, que destina una partida masiva de **124,5 millones de euros a subsidios mediáticos** bajo la premisa loable de apoyar la digitalización tecnológica del sector y

respaldar el periodismo de calidad, ha generado un intenso, polarizado y profundo debate crítico dentro de la profesión periodística.⁵³

Aunque la narrativa oficial y legislativa de la administración sostiene firmemente que el único objetivo de esta lluvia de millones es garantizar la viabilidad del tejido periodístico tradicional frente al embate desleal de las plataformas tecnológicas internacionales y dotar de herramientas a los medios para combatir las fábricas de desinformación automatizadas, analistas de medios independientes y comités de ética argumentan con preocupación que tal magnitud de financiación estatal condiciona severamente la ya maltrecha independencia editorial de las cabeceras. En el ecosistema hiper-competitivo de 2025, la pura subsistencia financiera y operativa de numerosos diarios regionales y nacionales, así como de emisoras de radio, depende ahora de manera directa, asimétrica y existencial de la complacencia, el favor y las partidas presupuestarias del poder ejecutivo gubernamental.⁵³

Esta gigantesca palanca de coerción económica se combina de manera estratégica con intensas presiones retóricas desde los púlpitos institucionales. La administración ha promovido en la sociedad una categorización discursiva fuertemente dicotómica y beligerante del ecosistema informativo. Esta táctica política establece una profunda división artificial entre los denominados medios "establecidos" (aliados estratégicos, respetables y, por ende, receptores legítimos de los vitales subsidios estatales) y una amalgama de nativos digitales a los que etiqueta peyorativamente como "pseudomedios".⁵³ Estos últimos son frecuentemente acusados de promover intereses extremistas de la derecha política, subvertir el orden democrático y esparcir bulos sistemáticamente mediante el uso ilegítimo de IA generativa.

Si bien es innegable que la desinformación estructurada, financiada de forma opaca y amplificadas algorítmicamente es una de las amenazas más corrosivas para la democracia occidental, el uso constante y punitivo de esta categorización estigmatizadora desde las esferas gubernamentales actúa, de facto, como un poderoso mecanismo disuasorio diseñado para silenciar, aislar financieramente o desacreditar sistemáticamente investigaciones periodísticas incómodas para el poder de turno. Esto es particularmente evidente y preocupante cuando dichas investigaciones documentan e interpelan casos de corrupción, mala praxis administrativa o sensibilidad judicial que afectan de manera directa a altos cargos institucionales, ministerios clave o incluso al entorno familiar inmediato de la presidencia del gobierno.⁵³

En esta misma línea de tensión político-mediática, la propuesta gubernamental de crear un registro obligatorio de medios de comunicación para forzar legalmente a las empresas a divulgar sus intrincadas fuentes de ingresos publicitarios y su estructura real de propiedad se inscribe en esta misma lógica de confrontación de poder: aunque oficialmente concebido y presentado ante la Unión Europea como una herramienta democrática de transparencia innegable, el registro es interpretado con profundo recelo por gran parte del sector periodístico independiente, que lo percibe como un posible

instrumento de escrutinio punitivo y control asimétrico contra las voces disidentes y los editoriales críticos.⁵³

La Captura Gubernamental de la Radiodifusión Pública y la Consiguiente Crisis de Confianza Ciudadana

Finalmente, el ecosistema de la radiodifusión pública, que teóricamente debería operar como el contrapeso institucional, imparcial y objetivo frente a las dinámicas mercantiles y la polarización de las plataformas privadas, representa en la práctica otro vértice crítico de esta condicionalidad política. Los abruptos cambios estructurales y legislativos implementados en la alta dirección y el consejo de la corporación pública Radio Televisión Española (RTVE) ilustran a la perfección la instrumentalización descarnada del ente público, financiado por los contribuyentes, en favor de estrategias políticas, partidistas y de manipulación de la opinión pública nacional a gran escala.⁵³

Mediante la imposición de un decreto ley extraordinario en 2024, el ejecutivo logró alterar profundamente las reglas históricas de gobernanza de la institución. La maniobra legislativa modificó las mayorías parlamentarias cualificadas que previamente eran imprescindibles para consensuar y nombrar a la cúpula directiva, reduciendo el requisito a una mera mayoría simple, permitiendo así sortear a la oposición política.⁵³ Paralelamente, la reforma incrementó el número de consejeros sentados en el consejo de administración, diluyendo voces disidentes, mientras que al mismo tiempo despojaba de competencias clave al propio consejo para concentrar poderes ejecutivos absolutos y sin precedentes en la figura del presidente de la corporación, afín al ejecutivo.⁵³ Con esta arquitectura institucional subyugada, el gobierno ha logrado alinear metódicamente la programación informativa de la cadena estatal con sus intereses narrativos y estratégicos a corto plazo.⁵³

El reflejo más ostensible y debatido de esta captura institucional fue la contratación multimillonaria, impulsada y protegida directamente desde las esferas gubernamentales, de figuras destacadas del entretenimiento puro, notablemente el fichaje por 12 millones de euros anuales de presentadores como David Broncano y su formato televisivo.⁵³ El propósito no declarado, pero ampliamente analizado por la crítica mediática, de estas monumentales inversiones de dinero público era competir de manera directa y destructiva en la franja horaria de máxima audiencia (prime time) contra los programas líderes de las cadenas de televisión privadas de línea editorial crítica hacia el gobierno, como el exitoso formato conducido por Pablo Motos. Este movimiento evidencia de forma incontestable cómo las decisiones corporativas de inversión, programación y foco de atención en los entes mediáticos, incluso en los de titularidad pública, obedecen a un calculado tablero de control político e influencia ideológica antes que a consideraciones estrictamente relacionadas con el servicio público, el rigor periodístico o la libertad de información.⁵³

El resultado sociológico combinado de la masiva manipulación algorítmica impulsada de

forma implacable por las plataformas tecnológicas globales, sumada a la profunda financiarización de los menguantes oligopolios mediáticos corporativos privados, y agravada por la intervención política estatal coercitiva de los gobiernos de turno, ha logrado erosionar de manera severa, y tal vez irreversible, el contrato social tácito que unía a la prensa institucional con la ciudadanía libre. El riguroso estudio empírico internacional plasmado en el *Digital News Report 2025* del Reuters Institute constata, con crudeza estadística, que la confianza de la sociedad en las noticias y los medios de comunicación en España se ha desplomado estrepitosamente a un mínimo histórico sin precedentes del 31%.⁵³ Este alarmante indicador sociopolítico confirma inequívocamente que el público general no solo sospecha de las falsificaciones generadas por IA, sino que percibe de manera palpable, profunda y fundamentada la constante instrumentalización, el sesgo algorítmico y la subordinación institucional de la información que consume diariamente.

Conclusiones sobre la Arquitectura del Control Informativo

La arquitectura global de la información pública se encuentra inmersa en una encrucijada existencial y crítica definida por la irrupción masiva, descentralizada y ubicua de la Inteligencia Artificial generativa multimodal. Los enormes ecosistemas digitales que monopolizan el tráfico web —Alphabet, Meta, TikTok, LinkedIn y X— han reconfigurado agresivamente los cimientos de sus algoritmos en un esfuerzo defensivo sin cuartel para evitar el colapso funcional de la utilidad de sus redes bajo el peso aplastante de la hiperinflación de medios sintéticos y spam industrializado.

Sin embargo, el presente análisis exhaustivo demuestra de manera concluyente que estas medidas técnicas de mitigación, desde las intrincadas redes neuronales de SpamBrain hasta la implementación incrustada de etiquetas de metadatos criptográficos inalterables como C2PA, no representan en modo alguno una neutralidad filantrópica, ni un compromiso incondicional con la verdad epistémica. Por el contrario, son contramedidas operativas intrínsecamente diseñadas para fortificar y proteger de la disrupción sus lucrativos modelos de negocio basados en la recolección masiva de la atención humana y la retención absoluta del tráfico para la explotación publicitaria segmentada.

Los datos recogidos de investigaciones académicas, reportes de la industria, marcos normativos europeos y casos de estudio nacionales documentan empíricamente que la dieta informativa diaria que asimila pasivamente la población mundial está fuertemente constreñida, filtrada y condicionada por múltiples capas superpuestas de intereses sistémicos, todos ellos completamente exógenos y contrarios a la salvaguarda del derecho fundamental a la libre información veraz:

1. **Condicionamiento Técnico-Algorítmico y Apropiación:** Los motores de recomendación, programados con lógicas capitalistas de plataforma, privilegian heurísticamente la emoción primaria, la indignación tribal y la polarización extrema ("shareworthiness") por encima del rigor empírico, la moderación analítica y la

objetividad periodística, todo ello para maximizar la métrica absoluta del tiempo en pantalla. A su vez, el despliegue parasitario de "AI Overviews" asfixia económicamente a los medios de comunicación originales al retener al usuario en entornos cerrados controlados de manera oligopólica por las corporaciones tecnológicas de Silicon Valley, estableciendo un "internet de cero clics" que vampiriza la creación humana.

2. **Condicionamiento Cognitivo y Manipulación de Identidad:** La exposición ininterrumpida a algoritmos deliberadamente opacos crea sofisticadas "burbujas generativas" e instituye sesgos socioculturales estructurales en la psique profunda del usuario final. Estas arquitecturas dificultan severamente la pluralidad, la fricción constructiva y la sanidad del discurso público a pesar de los loables esfuerzos por imponer etiquetas de advertencia, cuya eficacia real y comprobable sobre la alteración del comportamiento humano en la validación identitaria ha demostrado ser extremadamente limitada.
3. **Condicionamiento Político-Económico e Intervención Estatal:** Aterrizando el análisis en ecosistemas soberanos específicos como el español, la crisis sistémica de ingresos publicitarios, derivada de la disrupción provocada por la plataformización tecnológica global, ha dejado a la inmensa mayoría de los medios tradicionales al borde de la quiebra técnica, a merced de fondos de inversión de riesgo y atrapados por el intervencionismo del aparato estatal. La dependencia crítica de fondos y subsidios institucionales, la grosera instrumentalización estratégica de la radiodifusión pública para purgar narrativas inconvenientes, y la agresiva polarización retórica gubernamental en torno a la criminalización de los "pseudomedios", logran subordinar eficazmente la mermada independencia editorial periodística a la imperiosa necesidad de la supervivencia comercial corporativa y al apaciguamiento de la casta política dominante.

En última instancia, para restaurar la viabilidad y la pureza del derecho inalienable a la libre información, las sociedades democráticas avanzadas requieren trascender la ingenua y cosmética imposición legal de etiquetas estandarizadas sobre los modelos de IA. Se hace existencialmente indispensable la materialización de auditorías independientes y registros de algoritmos transparentes respaldados por entidades supervisoras fuertes con poderes sancionadores (como el prometido rol de la AESIA en España). Requiere, además, un fomento incansable, desde los estratos educativos más básicos, de una alfabetización mediática y tecnológica profundamente crítica que desarticule el consumo humano pasivo ante la vorágine de la economía de la atención.

Solamente a través de una reconfiguración ética, normativa y estructural del complejo entramado sociotécnico y político-financiero que gobierna el ecosistema mediático, la humanidad podrá garantizar que la intermediación algorítmica de la información y la revolución de la inteligencia artificial sirvan genuinamente a la emancipación cognitiva, al pensamiento crítico y a la madurez democrática de la sociedad civil, en lugar de perfeccionar la maquinaria de la explotación comercial, la manipulación epistémica y el

control sociopolítico.

Obras citadas

1. The Attention Economy Meets AI: A Marriage of Inconvenience ..., fecha de acceso: febrero 23, 2026, <https://www.nippon.com/en/in-depth/d01198/>
2. Google Search's guidance about AI-generated content, fecha de acceso: febrero 23, 2026, <https://developers.google.com/search/blog/2023/02/google-search-and-ai-content>
3. Vox Media's Editorial Ethics & Guidelines, fecha de acceso: febrero 23, 2026, <https://www.voxmedia.com/ethics/>
4. Algorithmic influence and media legitimacy: a systematic ... - Frontiers, fecha de acceso: febrero 23, 2026, <https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1667471/full>
5. AI Content & SEO in 2025: Can Google Detect AI-Generated Content? - RewriterPro.AI, fecha de acceso: febrero 23, 2026, <https://rewriterpro.ai/blog/can-google-detect-ai-generated-content>
6. How Google Detects Content Spam in 2025 - SearchX | SEO Agency, fecha de acceso: febrero 23, 2026, <https://searchxpro.com/how-google-detects-content-spam-in-2025/>
7. Google AI Search Updates and Spam Detection August 2025: What You Need to Know, fecha de acceso: febrero 23, 2026, <https://www.seoteric.com/google-ai-search-updates-and-spam-detection-august-2025-what-you-need-to-know/>
8. Why Lawsuits Over AI Summaries Will Fail: There is No Right to Traffic | TechPolicy.Press, fecha de acceso: febrero 23, 2026, <https://www.techpolicy.press/why-lawsuits-over-ai-summaries-will-fail/>
9. Google AI Overviews 2025: Top Cited Domains & Traffic Shifts - The Digital Bloom, fecha de acceso: febrero 23, 2026, <https://thedigitalbloom.com/learn/google-ai-overviews-top-cited-domains-2025/>
10. Google AI Overviews Reduce Publisher Traffic by 58%: Antitrust Analysis - ALM Corp, fecha de acceso: febrero 23, 2026, <https://almcorp.com/blog/google-ai-overviews-publisher-traffic-decline-antitrust-lawsuit-analysis/>
11. Google AI Overviews surged in 2025, then pulled back: Data - Search Engine Land, fecha de acceso: febrero 23, 2026, <https://searchengineland.com/google-ai-overviews-surge-pullback-data-466314>
12. Our Approach to Labeling AI-Generated Content and Manipulated ..., fecha de acceso: febrero 23, 2026, <https://about.fb.com/news/2024/04/metasp-approach-to-labeling-ai-generated-content-and-manipulated-media/>
13. Labeling AI-Generated Images on Facebook, Instagram and Threads, fecha de acceso: febrero 23, 2026, <https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/>
14. More Speech and Fewer Mistakes - About Meta | Social Technology, VR, AR, and Innovation, fecha de acceso: febrero 23, 2026,

- <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>
15. Meta unveils new rules for AI-generated content in political ads - FedScoop, fecha de acceso: febrero 23, 2026, <https://fedscoop.com/meta-unveils-new-rules-for-ai-generated-content-in-political-ads/>
 16. About AI-generated content - TikTok Support, fecha de acceso: febrero 23, 2026, <https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content>
 17. More ways to spot, shape and understand AI-generated content ..., fecha de acceso: febrero 23, 2026, <https://newsroom.tiktok.com/en-us/more-ways-to-spot-shape-and-understand-ai-generated-content>
 18. How the LinkedIn Algorithm Works [2024 Changes & Updates] - Tinuiti, fecha de acceso: febrero 23, 2026, <https://tinuiti.com/blog/paid-social/linkedin-algorithm/>
 19. LinkedIn Updates Tanking Reach - Allied Insight, fecha de acceso: febrero 23, 2026, <https://alliedinsight.com/resources/linkedin-updates-reach/>
 20. Understanding the LinkedIn Algorithm: Updates to Look Out for in 2024 - Bigg, fecha de acceso: febrero 23, 2026, <https://www.bigg.co.uk/learn/understanding-the-linkedin-algorithm-updates-to-look-out-for-in-2024>
 21. Commission investigates Grok and X's recommender systems under ..., fecha de acceso: febrero 23, 2026, https://ec.europa.eu/commission/presscorner/detail/en/ip_26_203
 22. The Policy Implications of Grok's 'Mass Digital Undressing Spree' | TechPolicy.Press, fecha de acceso: febrero 23, 2026, <https://www.techpolicy.press/the-policy-implications-of-groks-mass-digital-undressing-sprees/>
 23. California orders Elon Musk's AI company to immediately stop sharing sexual deepfakes, fecha de acceso: febrero 23, 2026, <https://calmatters.org/economy/technology/2026/01/california-investigates-deepfakes-elon-musk-company/>
 24. The Human Cost of Unregulated AI Tools, fecha de acceso: febrero 23, 2026, <https://www.hrw.org/news/2026/02/16/the-human-cost-of-unregulated-ai-tools>
 25. Labeling AI-generated media online - Adam J. Berinsky, fecha de acceso: febrero 23, 2026, https://berinsky.mit.edu/files/2026/01/labelingaigenerated_2025.pdf
 26. Trap of Social Media Algorithms: A Systematic Review of Research on Filter Bubbles, Echo Chambers, and Their Impact on Youth - MDPI, fecha de acceso: febrero 23, 2026, <https://www.mdpi.com/2075-4698/15/11/301>
 27. (PDF) Trap of Social Media Algorithms: A Systematic Review of Research on Filter Bubbles, Echo Chambers, and Their Impact on Youth - ResearchGate, fecha de acceso: febrero 23, 2026, https://www.researchgate.net/publication/397059811_Trap_of_Social_Media_Algorithms_A_Systematic_Review_of_Research_on_Filter_Bubbles_Echo_Chambers_and_Their_Impact_on_Youth
 28. Human or AI? Evaluating Labels on AI-Generated Social Media Content - The Dais, fecha de acceso: febrero 23, 2026, <https://dais.ca/reports/human-or-ai/>
 29. AI labeling reduces the perceived accuracy of online content but has limited broader effects We thank Verian UK Limited for providing the sample used in this study. We are grateful for the valuable feedback received at the APSA Annual

- Meeting (2024, 2025), particularly from Andrew Guess (discussant) and Leticia Bode (chair). We also appreciate the helpful suggestions from participants in seminars at the LSE Department of Methodology and the LSE Data Science Institute. Author contributions - arXiv, fecha de acceso: febrero 23, 2026, <https://arxiv.org/html/2506.16202v2>
30. Sesgo algorítmico: ¿Cómo la IA refleja sesgos humanos? - AI Automation, fecha de acceso: febrero 23, 2026, <https://aiautomation.com.co/etica-y-regulacion/sesgo-algoritmico-inteligencia-artificial-refleja-sesgos-humanos/>
 31. Inteligencia artificial y sesgos algorítmicos | Nueva Sociedad, fecha de acceso: febrero 23, 2026, <https://nuso.org/articulo/inteligencia-artificial-y-sesgos-algoritmicos/>
 32. Algorithmic Influence of Social Media Platforms on Public Opinion and Economic Decisions in the Digital Age - IDEAS/RePEc, fecha de acceso: febrero 23, 2026, <https://ideas.repec.org/a/abq/jirsd1/v4y2025i1p48-59.html>
 33. The generative AI bubble is changing how we see the world - LSE Business Review, fecha de acceso: febrero 23, 2026, <https://blogs.lse.ac.uk/businessreview/2025/03/28/the-generative-ai-bubble-is-changing-how-we-see-the-world/>
 34. Generative artificial intelligence-mediated confirmation bias in health information seeking - PMC, fecha de acceso: febrero 23, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12412720/>
 35. What the EU's New AI Code of Practice Means for Labeling Deepfakes | TechPolicy.Press, fecha de acceso: febrero 23, 2026, <https://www.techpolicy.press/what-the-eus-new-ai-code-of-practice-means-for-labeling-deepfakes/>
 36. Article 50: Transparency Obligations for Providers and Deployers of Certain AI Systems | EU Artificial Intelligence Act, fecha de acceso: febrero 23, 2026, <https://artificialintelligenceact.eu/article/50/>
 37. High-level summary of the AI Act | EU Artificial Intelligence Act, fecha de acceso: febrero 23, 2026, <https://artificialintelligenceact.eu/high-level-summary/>
 38. Code of Practice on marking and labelling of AI-generated content, fecha de acceso: febrero 23, 2026, <https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content>
 39. Estrategia Inteligencia Artificial 2024 - Ministerio de Economía, Comercio y Empresa, fecha de acceso: febrero 23, 2026, https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf
 40. la AESIA, fecha de acceso: febrero 23, 2026, <https://aesia.digital.gob.es/es>
 41. Un registro de algoritmos, el primer paso para una IA transparente en España - Civio, fecha de acceso: febrero 23, 2026, <https://civio.es/novedades/2025/03/14/registro-de-algoritmos-ia-transparencia-espana/>
 42. Registros de algoritmos para una transparencia real. - IA Ciudadana, fecha de acceso: febrero 23, 2026, <https://iaciudadana.org/wp-content/uploads/2025/03/Policy-Brief-ESP.pdf>
 43. El Congreso de los Diputados aprueba Ley General de Comunicación Audiovisual,

- fecha de acceso: febrero 23, 2026, <https://television.digital.gob.es/atencion-a-la-ciudadania/notas-prensa/ley-comunicacion-audiovisual>
44. Ley influencer en España: qué dice la nueva normativa y cómo afecta a las marcas, fecha de acceso: febrero 23, 2026, <https://prismalia.com/blog/ley-influencer-espana/>
 45. Algorithmic Attention and Content Creation on Social Media Platforms* - Economics, fecha de acceso: febrero 23, 2026, <https://www.econ.cuhk.edu.hk/wp-content/uploads/2025/11/YiChen.pdf>
 46. Transparencia algorítmica en los medios de comunicación y las ..., fecha de acceso: febrero 23, 2026, <https://www.revistatransparencia.com/ret/article/view/308>
 47. Facts: Google's push to AI hurts publisher traffic - Digital Content Next, fecha de acceso: febrero 23, 2026, <https://digitalcontentnext.org/blog/2025/08/14/facts-googles-push-to-ai-hurts-publisher-traffic/>
 48. Google AI Overviews linked to 25% drop in publisher referral traffic, new data shows, fecha de acceso: febrero 23, 2026, <https://digiday.com/media/google-ai-overviews-linked-to-25-drop-in-publisher-referral-traffic-new-data-shows/>
 49. fecha de acceso: febrero 23, 2026, <https://ppc.land/google-expands-internal-links-in-ai-overviews/#:~:text=Google%20has%20introduced%20a%20significant,made%20on%20April%2011%2C%202025.>
 50. Media Ownership and Transparency in Spain: An Analysis of Ownership Patterns and Regulatory Challenges - e-Repositori UPF, fecha de acceso: febrero 23, 2026, <https://repositori.upf.edu/bitstreams/384574ca-9aa9-4074-b0d6-f3f055fb7970/download>
 51. (PDF) Media Ownership and Transparency in Spain: An Analysis of Ownership Patterns and Regulatory Challenges - ResearchGate, fecha de acceso: febrero 23, 2026, https://www.researchgate.net/publication/392749783_Media_Ownership_and_Transparency_in_Spain_An_Analysis_of_Ownership_Patterns_and_Regulatory_Challenges
 52. Newspaper attention and policy activities in Spain | Journal of Public Policy, fecha de acceso: febrero 23, 2026, <https://www.cambridge.org/core/journals/journal-of-public-policy/article/newspaper-attention-and-policy-activities-in-spain/087C6017BF262160FD6D665E9EB91692>
 53. Spain | Reuters Institute for the Study of Journalism, fecha de acceso: febrero 23, 2026, <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/spain>
 54. Spain | Reuters Institute for the Study of Journalism, fecha de acceso: febrero 23, 2026, <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/spain>