

FORO DE COMUNICACIÓN

FUTURIBLE 2025

Gestión de la desinformación y
credibilidad en **MEDIOS DIGITALES**

Coordinadores

Octavio Islas
Amaia Arribas
Haydée Guzmán



 **Razón y Palabra**
Sello Editorial

Foro de Comunicación FUTURIBLE 2025

Gestión de la desinformación y credibilidad en medios digitales

Coordinadores:

Octavio Islas

Amaia Arribas

Haydée Guzmán



Título original: Foro de Comunicación FUTURIBLE 2025. Gestión de la desinformación y credibilidad en medios digitales.

2026 Octavio Islas, Amaia Arribas, Haydée Guzmán

Primera edición en Razón y Palabra, Sello Editorial, 2026

Diseño de tapas: Sabrina Sinatra

Diseño interior: Amaia Arribas Urrutia

ISBN: 978-9907-0-1255-2

Quito, Ecuador.

Sello Editorial Razón y Palabra <https://razonypalabraeditorial.com>

Los textos de esta obra han sido dictaminados por pares ciegos.



Este libro es una publicación de acceso abierto con los principios de Creative Commons Attribution 4.0 International License que permite el uso, intercambio, adaptación, distribución y transmisión en cualquier medio o formato, siempre que dé el crédito apropiado al autor, origen y fuente del material gráfico. Si el uso del material gráfico excede el uso permitido por la normativa legal deberá obtener el permiso directamente del titular de los derechos de autor.

Índice

PRÓLOGO

Octavio Islas Carmona y Amaia Arribas 6

Presentación de la primera jornada. *Haydée Guzmán* 11

MESA 1: GESTIÓN DE LA DESINFORMACIÓN Y CREDIBILIDAD EN LAS REDES SOCIALES 14

Moderación: Juan Carlos Peña Gutiérrez. BOLIVIA

Capítulo 1

**Sesgos, desinformación y la era de la inteligencia artificial: un
enfoque estratégico para líderes empresariales** 17

Antonio Salgado Leiner. MÉXICO

Capítulo 2

**La reputación organizacional en la era digital: un relato de
sombras y resiliencia** 30

Alejandro Salum. MÉXICO

MESA 2: GESTIÓN DE LA DESINFORMACIÓN Y CREDIBILIDAD EN LA POLÍTICA Y LA GOBERNANZA 41

Moderación: María Eugenia Garcés. ECUADOR

Capítulo 3

**El impacto de gobernar en medio de emociones y narrativas
virales** 44

Gabriel Salazar Navarro. COLOMBIA

Presentación de la segunda jornada. *Dona Arrieta* 65

MESA 3: GESTIÓN DE LA DESINFORMACIÓN Y CREDIBILIDAD EN LA INTELIGENCIA ARTIFICIAL 67

Moderación: Maritza Mosquera de Sumich. PANAMÁ

Capítulo 4

IA y desinformación como catalizadores de violencia en línea 70

Jorge Rendón Riba. ESPAÑA

Capítulo 5

**La verdad bajo amenaza: desinformación y contenidos sintéticos
en una nueva ecología informativa** 83

Soledad Arréquez Manozzo. ARGENTINA

PRÓLOGO

Estrategia, desinformación y verosimilitud en la era de la complejidad

Para comprender la densidad analítica y el alcance conceptual de los trabajos reunidos en esta obra, resulta imprescindible detenerse en el marco teórico que los articula: la Nueva Teoría Estratégica (NTE). No estamos ante una simple actualización terminológica ni frente a un ejercicio de reformulación superficial de modelos preexistentes. La NTE representa, en sentido estricto, un cambio paradigmático en la forma de entender la estrategia, la comunicación y, de manera muy especial, los fenómenos contemporáneos de la desinformación.

6

Durante gran parte del siglo XX, la estrategia fue concebida desde una lógica instrumental, profundamente marcada por el pensamiento cartesiano y mecanicista. En ese paradigma clásico, la estrategia equivalía a planificación racional orientada al control: un marco estable, actores claramente delimitados, objetivos definidos con precisión y un conjunto de acciones diseñadas para maximizar la eficiencia y minimizar la incertidumbre. La metáfora del ajedrez sintetiza de manera elocuente esta concepción: un tablero fijo, piezas con roles estables, reglas claras y un adversario cuya derrota constituía el objetivo último.

Sin embargo, la realidad comunicativa del siglo XXI ha erosionado de forma irreversible los supuestos sobre los que se asentaba ese modelo. El desarrollo de las plataformas digitales, la hiperconectividad permanente, la fragmentación de las audiencias y la automatización algorítmica han dado lugar a un ecosistema relacional fluido, caracterizado por la inestabilidad, la no linealidad y la imprevisibilidad. En este nuevo escenario, el tablero ha desaparecido, las piezas ya no permanecen bajo control y los movimientos generan efectos colaterales imposibles de prever desde lógicas deterministas.

Es precisamente en este contexto de complejidad donde la Nueva Teoría Estratégica –impulsada por autores como Rafael Alberto Pérez y desarrollada de manera colectiva por académicos y profesionales, entre ellos los participantes en el Foro Futurable 2025– adquiere toda su relevancia. La NTE propone abandonar la obsesión por el control y la eficiencia instrumental para comprender la estrategia como una ciencia de la relación, orientada a gestionar dinámicas complejas de sentido, confianza e interacción social.

El pensamiento estratégico tradicional descansaba sobre una serie de supuestos que, durante décadas, parecieron incuestionables:

- La existencia de entornos relativamente estables.
- La identificabilidad clara de los actores estratégicos.
- La racionalidad de los sujetos implicados.
- La posibilidad de prever los efectos de las acciones implementadas.
- La centralidad del control como objetivo último.

Estos supuestos permitieron desarrollar modelos eficaces en contextos industriales, militares o corporativos relativamente cerrados. Sin embargo, la transformación digital ha puesto de manifiesto la fragilidad de estos fundamentos. En entornos abiertos, hiperconectados y mediáticamente saturados, las lógicas lineales dejan de ser operativas.

La comunicación ya no circula de manera unidireccional ni obedece a jerarquías claras. Los mensajes se reinterpretan, se remezclan y se resignifican en tiempo real, generando dinámicas emergentes que escapan al control de los emisores originales. En este contexto, insistir en modelos estratégicos diseñados para entornos predecibles no solo resulta ineficaz, sino potencialmente contraproducente. La NTE parte del reconocimiento explícito de la complejidad como condición estructural de la comunicación contemporánea. Esta complejidad se manifiesta en múltiples dimensiones:

- Complejidad tecnológica, derivada de la interacción entre plataformas, algoritmos y sistemas automatizados.
- Complejidad social, vinculada a la diversidad cultural, ideológica y emocional de las audiencias.
- Complejidad temporal, marcada por la aceleración y la simultaneidad.
- Complejidad semántica, producto de la proliferación de narrativas, interpretaciones y marcos de sentido.

Asumir la complejidad no implica renunciar a la estrategia, sino redefinirla. Desde la NTE, la estrategia deja de ser un plan cerrado para convertirse en un proceso adaptativo, abierto y en permanente ajuste, orientado a aprender del entorno relacional y a intervenir en él de manera reflexiva.

Uno de los aportes centrales de la Nueva Teoría Estratégica es el desplazamiento del foco estratégico desde la transmisión de mensajes hacia la trama relacional en la que dichos mensajes adquieren significado. Frente al modelo clásico —emisor, mensaje, receptor—, la NTE concibe la comunicación como un fenómeno relacional y emergente, en el que el sentido se construye en la interacción.

Desde esta perspectiva, los mensajes no poseen un significado fijo que se transfiere intacto de un punto a otro. Por el contrario, su sentido se produce en contextos específicos, atravesados por emociones, identidades, expectativas y relaciones de poder. Este cambio de enfoque resulta especialmente relevante para comprender la dinámica de la desinformación. La proliferación de contenidos falsos o engañosos no puede explicarse únicamente por déficits informativos ni por fallos en los mecanismos de verificación. En muchos casos, la información falsa circula, se reproduce y se legitima no por su veracidad empírica, sino por su capacidad para reforzar vínculos, activar emociones compartidas y consolidar identidades colectivas.

Desde esta óptica, la disputa en torno a la desinformación ya no se libra únicamente en el terreno del dato, sino en el del vínculo. Difundir información verdadera, por sí sola, resulta insuficiente cuando el tejido relacional entre una institución y su audiencia se encuentra deteriorado. Una organización que ha perdido credibilidad no la recuperará mediante estadísticas o comunicados si antes no reconstruye la relación de confianza que da sentido a esos mensajes.

8

La NTE pone de relieve que la desinformación florece especialmente en contextos donde los vínculos están erosionados:

- Percepciones de abandono o exclusión.
- Sentimientos de agravio o invisibilización.
- Crisis de legitimidad institucional.
- Déficits de reconocimiento simbólico.

En estos escenarios, los relatos falsos no solo desinforman, sino que reconectan emocionalmente a individuos con comunidades imaginadas que les ofrecen pertenencia, sentido y reconocimiento. Otro de los pilares conceptuales de la Nueva Teoría Estratégica es la revisión crítica del modelo antropológico implícito en la estrategia tradicional. Durante décadas, se asumió la existencia de un sujeto racional, capaz de procesar información objetivamente y de tomar decisiones óptimas cuando los datos eran claros y accesibles. La experiencia contemporánea de la desinformación ha desmentido esta premisa de forma contundente. Numerosos estudios han mostrado que la exposición a información verificada no garantiza la corrección de creencias falsas. En determinados contextos, la refutación factual puede incluso reforzar la adhesión a narrativas erróneas, fenómeno conocido como *backfire effect*.

Frente a esta imagen reduccionista, la NTE propone la figura del sujeto relacional: un individuo inscrito en redes sociales, culturales y emocionales, cuyo comportamiento comunicativo está profundamente condi-

cionado por su pertenencia a comunidades simbólicas. En este marco, la desinformación no opera únicamente sobre déficits cognitivos, sino que moviliza emociones, sesgos y mecanismos identitarios. La comunicación se convierte así en un campo de batalla simbólico donde compiten distintos regímenes de verdad.

La llamada “verdad emocional” —aquella que refuerza identidades y ofrece coherencia narrativa— puede imponerse con facilidad a la verdad científica cuando esta última carece de anclaje relacional. Comprender este desplazamiento resulta esencial para diseñar estrategias eficaces frente a la desinformación. Uno de los aportes más provocadores de la NTE es la aceptación explícita de que el control absoluto del flujo informativo es una ilusión. En entornos dominados por plataformas digitales, algoritmos de recomendación y procesos de amplificación automatizada, ninguna entidad puede aspirar a gobernar de manera total la circulación de la información.

En consecuencia, la estrategia ya no puede concebirse como un plan cerrado ejecutado de forma lineal. Desde la NTE, la estrategia se redefine como un proceso:

- Abierto.
- Flexible.
- Iterativo.
- Basado en el aprendizaje continuo.

Gestionar la complejidad implica asumir la incertidumbre como condición estructural y desarrollar capacidades de observación, escucha y ajuste permanente. Si la NTE ofrece el marco conceptual para interpretar la desinformación, la inteligencia artificial generativa (IAG) constituye uno de los desafíos técnicos más disruptivos abordados en esta obra. Tal como se debatió en el Foro Futurible 2025, asistimos a un tránsito histórico: de una era centrada en la edición de la realidad a otra basada en su generación artificial.

Durante siglos, la confianza humana se apoyó en un principio aparentemente inquebrantable: ver para creer. La fotografía y el vídeo funcionaron como certificadores de autenticidad. La IA generativa ha quebrado ese contrato ontológico al permitir la creación de imágenes, audios y vídeos sintéticos prácticamente indistinguibles de los reales. Esta erosión no solo facilita la producción de noticias falsas, sino que introduce un fenómeno aún más corrosivo: el dividendo del mentiroso. Cuando cualquier evidencia puede ser desacreditada bajo la sospecha de haber sido generada artificialmente, la desinformación no solo induce a creer lo falso, sino que impide creer lo verdadero.

Los sistemas algorítmicos no se limitan a distribuir información: configuran activamente los entornos de visibilidad donde se forman opiniones y se consolidan creencias. Diseñados para maximizar métricas de atención, tienden a priorizar contenidos emocionalmente intensos y polarizantes, actuando como aceleradores de la desinformación y generando cámaras de eco.

Desde la NTE, la estrategia debe intervenir —en la medida de lo posible— en las condiciones estructurales de circulación del sentido, reconociendo que la visibilidad se ha convertido en un recurso estratégico central. Lejos de cualquier catastrofismo tecnológico, el tramo final de la obra adopta una mirada prospectiva orientada a la gobernanza de la complejidad. Mirar hacia el horizonte 2030–2040 implica abandonar una lógica reactiva y avanzar hacia una lógica anticipatoria. Desde la NTE, esta transición se articula en tres grandes líneas de acción:

- Alfabetización mediática de nueva generación.
- Sistemas de trazabilidad y verificación de origen.
- Ética tecnológica incorporada desde el diseño.

10

Para finalizar, la tesis central de esta obra es tan incómoda como necesaria: la crisis de la desinformación no es, en esencia, una crisis de datos, sino una crisis de relaciones. En sociedades complejas, la verdad no se impone por mera evidencia empírica, sino que se construye y se sostiene dentro de tramas relacionales vivas.

La Nueva Teoría Estratégica no ofrece soluciones simples. Ofrece, en cambio, un marco robusto para comprender por qué la verdad se debilita cuando se rompen los vínculos y por qué solo puede sobrevivir allí donde existen confianza, reconocimiento mutuo y sentido compartido.

Octavio Islas Carmona

Doctor en Ciencias Sociales. Catedrático e investigador en la Universidad Autónoma de Baja California (México), y Universidad Internacional de Andalucía (España). Director de la Cátedra Itinerante de la Nueva Teoría Estratégica (CINTE).

jose.islas.carmona@uabc.edu.mx

Amaia Arribas Urrutia

Doctora en Ciencias de la Información. Docente e investigadora en la Universidad de Valladolid (España). Presidenta del Capítulo España de FISEC.

amaya.arribas@uva.es

Presentación de la primera jornada

Por sexto año consecutivo, los miembros de FISEC y de CINTE entregamos FUTURIBLE, evento que nos permite encontrarnos para compartir diferentes aspectos importantes, de actualidad y de interés para la comunidad académica de diversas disciplinas en Iberoamérica.

Elegir un tema en el amplio espectro de nuestros campos de conocimiento y con la cantidad de proposiciones, asuntos, perspectivas, enfoques de hoy en día no es tarea fácil. Sin embargo, vimos que la Gestión de la desinformación y credibilidad en medios digitales, es un tema global, de alto impacto y sobre el cual, a pesar de que lo vivimos diariamente y lo analizamos, aún tenemos mucho que aprender.

La particularidad del momento actual se llama digitalización. La tecnología digital no es simplemente una tecnología de la comunicación, es una tecnología de la vida; es decir, nada escapa al mundo digital. Lo virtual compite con lo real y terminan por hibridarse.

El desarrollo digital propicia múltiples avances y beneficios, pero también un nuevo contexto social e informativo que ha favorecido una amplificación sin precedentes, de la desinformación y sus efectos, convirtiéndose realmente en una amenaza para personas y sistemas. Una campaña de desinformación provoca falsas creencias, confusión, desconfianza, divide y amplifica sesgos y prejuicios. Y se ve favorecida porque cualquier persona puede generar contenido, difundirlo y compartirlo lo que lleva a una explosión informativa que dificulta la identificación de contenidos veraces provocando incertidumbre.

11

La Desinformación se define entonces como información falsa, imprecisa o engañosa que se fabrica y se difunde con la intención de confundir, influir o manipular a la audiencia. Esta información puede ser deliberadamente errónea, o puede ser información verdadera que se presenta de una manera engañosa o se saca de contexto para generar una falsa impresión. El rasgo distintivo de la Desinformación está, entonces, en la intencionalidad deliberada de generarla o difundirla para provocar confusión o manipulación, y en la dificultad de verificarla de manera fáctica o documental.

Así, el primer tema que nos ocupa hoy es la Gestión de la desinformación y credibilidad en los medios de comunicación y en las redes digitales.... ¿Qué decir al respecto?

La coexistencia de medios digitales y tradicionales es esencial para una sociedad informada, pero la revolución digital ha transformado la forma

en que consumimos noticias. Y ¿cuál es el impacto que ha tenido en la credibilidad de los medios tradicionales? Y qué no decir en las redes sociales... Pero también es cierto que la desinformación en redes sociales ha contribuido a un escepticismo generalizado hacia los medios.

Esto plantea una amenaza no solo para la integridad del periodismo sino también para la confianza del público en los medios. Los consumidores de noticias deben aprender a discernir entre fuentes confiables y aquellas que pueden tener agendas ocultas. La evolución tecnológica no debe comprometer la calidad y la veracidad de la información que consumimos y compartimos.

El desafío de mantener la credibilidad en un mundo digitalizado es un reto crucial para los medios de comunicación tradicionales y para las redes sociales, pero también es una responsabilidad compartida entre estos, la tecnología y el público.

12 En cuanto a la Gestión de la desinformación y credibilidad en la política y la gobernanza diremos que la Desinformación, especialmente en su dimensión política, no es un fenómeno nuevo. La elaboración y difusión de mensajes que sobredimensionan rasgos de gobernantes, dirigentes, líderes de oposición, adversarios políticos locales o internacionales es natural e inherente a las disputas por el Poder que tienen un fin específico: obtener una ventaja política, económica o social a expensas de la verdad; o bien, producir, de manera deshonestamente, una desventaja para la reputación o Legitimidad de alguna acción, política o actor público. Quizás la única certeza que tenemos hoy en día es que la política está siendo muy diferente a como la conocimos en el pasado, y en el pasado reciente. No sabemos todavía si las nuevas tecnologías la mejorarán o la harán aún más difícil, pero es un hecho que ya están impactando nuestra vida política y modificando sus dinámicas.

Hoy, las plataformas, al tiempo que permiten una conversación, la condicionan de diversos modos. Por ejemplo, durante unas elecciones, la IA puede utilizarse eficazmente para implicar a los votantes a nivel individual a lo largo de todo el proceso electoral. Los chatbots y los foros de debate en las plataformas de redes sociales, que animan a las personas a dejar sus comentarios al final, son varias formas de medir el estado de ánimo del público. Además, puede ayudar a recopilar todos estos datos en tiempo real y permitir a las personas responsables de las campañas de los partidos ir cambiando de rumbo en función de la opinión del público. Y puede utilizarse para manipular a votantes. Esto ha sido evidente en este año electoral para varios países latinoamericanos y lo estamos viviendo ahora en Colombia.

Las valiosas aportaciones de quienes se dedican al *fact checking* no necesariamente significan que las evidencias se impondrán y dejarán de emitirse las falsedades en el espacio digital. Entre otras cosas, ésta limitada eficacia del *fact checking* se debe a que la conversación colectiva se refiere solo en una pequeña parte a objetividades y en una mayor medida al modo como nosotros interpretamos la realidad.

Y, por otra parte, La Gestión de la desinformación y credibilidad en la Inteligencia Artificial nos lleva a pensar que su desarrollo, aunque intensifica el alcance y peligro que supone, también ofrece nuevas oportunidades para detectarla y combatirla.

Herramientas basadas en IA están ayudando a detectar noticias falsas, pero también pueden ser manipuladas. La carrera por controlar “la verdad” digital se intensifica. Algunos medios ya experimentan con contenido generado completamente por IA. Esto puede erosionar la confianza si el público no sabe distinguir entre lo auténtico y lo automatizado.

Por lo tanto, la educación digital, la alfabetización mediática, el pensamiento crítico, el uso de tecnología para detectar contenido falso, el compromiso generalizado para difundir información verificada y corregir falsedades son solo algunos aspectos para lograr un entorno digital más seguro y responsable.

13

Pero dejemos que sean los expertos, a quienes agradecemos su participación, quienes nos compartan su conocimiento y experiencia sobre los temas previstos.

Muchas gracias.

Haydée Guzmán

Vicepresidenta de FISEC
guzmanramirezhaydee@gmail.com

MESA

Gestión de la desinformación y la credibilidad en las redes sociales

1

14

Moderador

Juan Carlos Peña Gutiérrez

Comunicólogo, académico y conferencista con formación en Comunicación Social y Dirección de Marketing. Embajador NEO del Sistema Original de Comunicación Estratégica®.

Cuenta con estudios de posgrado en educación superior, investigación científica, imagen corporativa y producción audiovisual. Inició su trayectoria docente universitaria en 1991 en Santa Cruz de la Sierra, Bolivia.

Ha impartido clases en universidades como UPSA, UPDS y UEB hasta 2017.

Desde 2017 es Jefe de la carrera de Comunicación Estratégica y Digital en UTEPSA. Posee amplia experiencia como consultor en diagnóstico y planificación comunicacional.

Especialista en gestión de prensa, imagen de marca, crisis y comunicación digital.

Ha ocupado cargos directivos en organizaciones empresariales y académicas.

Autor y coautor de diversos libros e investigaciones, con publicaciones en medios y revistas especializadas. jefe.carrera.comunicacion@utepsa.edu

La Conferencia FUTURIBLES 2025 se sumergió en uno de los temas más complejos y urgentes del mundo digital que es la desinformación y su efecto en la credibilidad pública. En un entorno donde las redes sociales y los medios digitales dominan, la propagación de información falsa ha pasado de ser algo excepcional a convertirse en una dinámica habitual que impacta tanto en las decisiones individuales como en la reputación de organizaciones e instituciones.

Tuve la oportunidad de moderar las conferencias de Antonio Salgado Leiner y Alejandro Salum, que dentro del tema “Gestión de la desinformación y credibilidad en los medios de comunicación y las redes sociales”, quienes analizaron estos temas desde la óptica profesional y académica.

Al inicio, Antonio Salgado Leiner enfatizó que, aunque la desinformación no es un fenómeno nuevo, su alcance actual se ve amplificado por la rapidez y la magnitud de los entornos digitales, de esta forma hizo hincapié en la necesidad de diferenciar entre información errónea, desinformación y malinformación, dependiendo de la intención detrás de la difusión del contenido. Al mismo tiempo, explicó cómo nuestros sesgos cognitivos, como el sesgo de confirmación o la heurística de disponibilidad, nos hacen más susceptibles a aceptar información falsa, especialmente en un entorno saturado de estímulos y marcado por la inmediatez. Además, advirtió sobre el impacto de tecnologías como la inteligencia artificial generativa, los bots y los deepfakes, que agravan el problema al generar contenidos que parecen muy creíbles. Ante este panorama, propuso una respuesta integral que combine alfabetización mediática, regulación, herramientas tecnológicas y colaboración social.

15

Por otro lado, Alejandro Salum examinó el impacto reputacional y económico de la desinformación en las organizaciones, subrayando que los ciberataques, la filtración de datos y las campañas de desinformación son amenazas constantes, con pérdidas globales proyectadas de 9.8 billones de dólares para 2025. Sin embargo, el daño va más allá de lo financiero pues afecta la confianza del público, la percepción de marca y la estabilidad institucional, destacando el papel de las redes sociales como amplificadoras de crisis y la necesidad de una preparación estratégica que incluya monitoreo, vocerías capacitadas y comunicación efectiva.

Ambas ponencias llegan a una conclusión fundamental, indicando que la desinformación es un fenómeno complejo que no se puede abordar únicamente desde un enfoque técnico, debido a que requiere un pensamiento crítico, una ética pública y una corresponsabilidad en-

tre los diferentes actores sociales, gubernamentales y tecnológicos. En este nuevo panorama, la credibilidad se convierte en un recurso estratégico que necesita ser construido y protegido de manera activa.

Sesgos, desinformación y la era de la inteligencia artificial: un enfoque estratégico para líderes empresariales

17

Antonio Salgado Leiner

Doctor en ciencias de la computación con especialidad en tecnología multimedia, por la Universidad Estatal de Minnesota donde también obtuvo su doctorado en Psicología experimental. Es Maestro en psicoacústica. Igualmente, Maestro en psicología evolutiva por la universidad de California en Los Ángeles. Interesado en temas de economía del comportamiento y las neurociencias aplicadas a la comedia y a la magia. Subraya que es músico, Standupero (comediante) amante del café y de la ciencia ficción, Ha colaborado en proyectos de comunicación digital, tecnología creativa para Marketing, sistemas de automatización de procesos administrativos y de marketing. Es consultor digital en grandes empresas multinacionales. Docente y tallerista en reconocidas instituciones de educación superior. asalgado@neosocialmedia.com

Introducción: La Encrucijada de la información y la cognición en la era digital

La proliferación de la información en la era digital ha traído consigo beneficios sin precedentes, pero también desafíos complejos, siendo la desinformación uno de los más acuciantes. Este fenómeno, caracterizado por la difusión intencionada o no de información falsa o engañosa, se ve amplificado exponencialmente por la velocidad y el alcance de las redes sociales y, más recientemente, por la capacidad generativa de la Inteligencia Artificial (IA). La IA, en particular los Grandes Modelos de Lenguaje (LLMs), ha transformado la creación y diseminación de contenido, haciendo que la distinción entre lo real y lo fabricado sea cada vez más difusa.

Es crucial reconocer que la desinformación no es un fenómeno exclusivo de la era digital. A lo largo de la historia, diversas formas de comunicación han sido utilizadas para manipular la percepción pública y difundir narrativas falsas. Ejemplos notables incluyen la transmisión radiofónica de “La Guerra de los Mundos” en Estados Unidos (1938) y Ecuador (1949), que generaron pánico masivo al ser percibidas como noticias reales (Cantril, 1940, Radio Ambulante, s.f.). De manera más sistemática y maliciosa, las campañas de propaganda de Joseph Goebbels en la Alemania nazi utilizaron la radio, el cine y la prensa para difundir desinformación y justificar atrocidades a gran escala (Welch, 1993). Estos precedentes históricos subrayan que la desinformación explota vulnerabilidades humanas y sociales que trascienden la tecnología.

Este artículo se adentra en la naturaleza de la desinformación desde una perspectiva multidisciplinar, examinando los fundamentos psicológicos de los sesgos humanos, la particular susceptibilidad de ciertos grupos demográficos, y proponiendo un enfoque estratégico para que las organizaciones, especialmente sus líderes, aborden la cuestión de los sesgos en el entrenamiento de los LLMs. Integrando los principios de la Nueva Teoría Estratégica (NTE), se busca ofrecer una hoja de ruta para construir ecosistemas digitales más inmunes a la desinformación, donde la confianza y la veracidad sean pilares fundamentales.

La Psicología de los sesgos y la desinformación: una perspectiva cognitiva y social

La desinformación no prospera en un vacío; se ancla y se propaga explotando las características inherentes de la cognición humana y las

dinámicas sociales. Comprender los sesgos psicológicos es fundamental para desentrañar por qué la información errónea es tan persuasiva y resistente a la corrección.

Sesgos cognitivos: atajos mentales explotados por la desinformación

Los sesgos cognitivos son atajos mentales (heurísticas) que nuestro cerebro utiliza para procesar información de manera rápida y eficiente. Si bien son adaptativos en muchas situaciones, también nos hacen vulnerables a la desinformación (Mschools, s.f.).

- Sesgo de confirmación: Quizás el sesgo más potente en el contexto de la desinformación. Es la tendencia a buscar, interpretar y recordar información de una manera que confirma nuestras creencias pre-existentes, mientras se ignora o se desacredita la información que las contradice (Blanco et al., 2024). En un entorno digital donde los algoritmos nos exponen a contenido que ya se alinea con nuestras opiniones, el sesgo de confirmación se ve amplificado, creando “cámaras de eco” y “burbujas de filtro” que refuerzan la desinformación.

19

- Heurística de disponibilidad: Tendemos a sobreestimar la probabilidad de eventos o la veracidad de la información que es más fácil de recordar o que viene a nuestra mente con mayor facilidad (YLA Network, 2019). La desinformación, a menudo diseñada para ser emocionalmente impactante o repetitiva, se vuelve más “disponible” en nuestra memoria, lo que aumenta su percepción de veracidad.

- Efecto de ilusión de verdad (fluidez de procesamiento): La exposición repetida a una afirmación, incluso si es falsa, aumenta la facilidad con la que se procesa (fluidez de procesamiento), lo que a su vez incrementa la probabilidad de que sea percibida como verdadera (The Decision Lab (s.f.). Las campañas de desinformación a menudo se basan en la repetición constante de narrativas falsas para explotar este efecto.

- Sesgo de anclaje: La tendencia a depender demasiado de la primera pieza de información ofrecida (el “ancla”) al tomar decisiones. Una vez que una narrativa falsa se establece como ancla, es difícil desplazarla, incluso con evidencia contradictoria (Área Humana (s.f.))

- Efecto *Dunning-Kruger*: Un sesgo cognitivo en el que individuos con poca habilidad o conocimiento en un área tienden a sobreestimar su propia competencia, mientras que aquellos con alta competencia tienden a subestimarla (Daría, 2024). Esto puede llevar a que personas con conocimientos limitados sobre un tema se sientan lo suficiente-

mente informadas como para discernir la verdad, haciéndolas más propensas a aceptar y compartir desinformación.

La Dimensión social de la desinformación

La desinformación no es solo un fenómeno cognitivo individual; es profundamente social. La forma en que interactuamos con otros y la confianza que depositamos en nuestras redes influyen significativamente en su propagación.

- Confianza social y redes de afinidad: Las personas son más propensas a creer y compartir información que proviene de fuentes en las que confían, como amigos, familiares o miembros de su comunidad (AARP Blogs, 2022). En un entorno digital donde los algoritmos nos exponen a contenido que ya se alinea con nuestras opiniones, el sesgo de confirmación se ve amplificado, creando “cámaras de eco” y “burbujas de filtro” que refuerzan la desinformación.

20 - Presión social y conformidad: La necesidad de pertenecer y conformarse a las normas del grupo puede llevar a los individuos a aceptar y compartir información, incluso si tienen dudas sobre su veracidad, para evitar el desacuerdo o la exclusión social (Montiel Torres et al., 2025). Esto es particularmente relevante en grupos polarizados donde la identidad grupal se construye en torno a narrativas específicas, verdaderas o falsas.

- Polarización afectiva: La desinformación a menudo se utiliza para exacerbar la polarización afectiva, es decir, la tendencia a ver a los miembros del propio grupo de manera positiva y a los del grupo opuesto de manera negativa (Pérez Colomé, 2025). Al demonizar al “otro” con información falsa, se refuerza la cohesión interna del grupo y se justifica la hostilidad hacia los externos.

La vulnerabilidad de los adultos mayores (45-55+ años) a la desinformación

Un aspecto crítico en la propagación de la desinformación es la aparente mayor susceptibilidad de las personas mayores de 45 o 55 años a compartir información errónea sin una verificación adecuada. Esta vulnerabilidad no es una deficiencia inherente a la edad, sino el resultado de una compleja interacción de factores cognitivos, sociales y tecnológicos (Parra, 2022).

Factores cognitivos

Si bien la inteligencia cristalizada (conocimiento acumulado) se mantiene o incluso mejora con la edad, la inteligencia fluida (velocidad de procesamiento, memoria de trabajo, flexibilidad cognitiva) puede experimentar cambios sutiles que impactan la interacción con la desinformación:

- Cambios en la memoria de trabajo y la inhibición: Una ligera disminución en la memoria de trabajo y en la capacidad de inhibir información irrelevante o distractora puede dificultar el procesamiento profundo y crítico de múltiples fuentes de información. Esto puede llevar a una mayor dependencia de la primera impresión o de la información más fácilmente accesible (Sádaba et al., 2023).

- Mayor susceptibilidad al sesgo de familiaridad: La exposición repetida a la desinformación, común en las redes sociales, puede hacer que la información falsa parezca más familiar y, por ende, más creíble, explotando el efecto de ilusión de verdad. Los adultos mayores, al pasar más tiempo en ciertas plataformas o grupos, pueden estar más expuestos a esta repetición (Pérez Colomé, 2025).

- Menor escepticismo hacia fuentes digitales novedosas: A diferencia de las generaciones más jóvenes que han crecido en un entorno digital saturado de información y, por lo tanto, han desarrollado un escepticismo innato hacia las fuentes en línea, algunos adultos mayores pueden no haber desarrollado las mismas habilidades de discernimiento para el ecosistema digital. Su experiencia previa con medios tradicionales (prensa, radio, TV) donde la verificación editorial era más rigurosa, puede llevarlos a asumir un nivel de veracidad similar en el entorno digital (Scherer & Pennycook, 2021).

Factores sociales y motivacionales

Las dinámicas sociales y las motivaciones personales juegan un papel significativo en la propensión a compartir desinformación:

- Búsqueda de conexión social y pertenencia: Las redes sociales ofrecen una vía vital para la conexión social, especialmente para aquellos que pueden experimentar aislamiento. Compartir contenido, incluso si es desinformación, puede ser una forma de mantener el contacto, sentirse parte de una comunidad y recibir validación social. La presión de grupo dentro de círculos cerrados (familiares, amigos, grupos de interés) puede incentivar el compartir sin verificar (AARP Blogs, 2022).

- Confianza en redes cercanas: Los adultos mayores tienden a confiar más en la información que proviene de sus redes sociales cercanas (familiares, amigos, conocidos) (The Conversation, 2020). Si estas personas comparten desinformación, es más probable que la acepten como verdadera, independientemente de la fuente original. Esta confianza, construida a lo largo de años, puede ser un vector poderoso para la desinformación.

- Refuerzo de identidad y afiliación ideológica: La desinformación a menudo se alinea con identidades políticas o ideológicas preexistentes. Los adultos mayores, que a menudo tienen identidades más consolidadas, pueden ser particularmente susceptibles a la desinformación que refuerza sus afiliaciones y demoniza a los grupos opuestos. Compartir esta información puede ser una forma de expresar su identidad y solidaridad con su grupo (Montiel Torres et al., 2025).

Factores tecnológicos y de alfabetización digital

22

La brecha digital, aunque decreciente, sigue siendo un factor relevante:

- Menor familiaridad con las dinámicas de las plataformas: Muchos adultos mayores no crecieron con internet y las redes sociales, lo que puede llevar a una menor comprensión de cómo funcionan los algoritmos, la viralidad de la información, la existencia de bots o cuentas falsas, y las tácticas de manipulación digital. Pueden no estar tan familiarizados con las señales de alerta de la desinformación en línea (Rivas De La Rosa, 2022).

- Diseño de plataformas: Las plataformas están diseñadas para maximizar el *engagement*, a menudo priorizando el contenido emocionalmente cargado o sensacionalista, que es precisamente el tipo de contenido que suele ser desinformación. Aunque esto afecta a todos, los adultos mayores pueden ser menos conscientes de cómo estos diseños influyen en su consumo de información.

- Brecha de habilidades de verificación específicas: Las habilidades de alfabetización mediática tradicionales no siempre se traducen directamente a la verificación de información en línea. Pueden carecer de las habilidades específicas para realizar búsquedas inversas de imágenes, verificar fuentes en línea o identificar sitios web de noticias falsas (Hernández, s.f.).

Estrategias para líderes empresariales: mitigando sesgos en el entrenamiento de LLMs desde la Nueva Teoría Estratégica

La creciente adopción de Modelos de Lenguaje Grandes (LLMs) en el ámbito empresarial presenta una oportunidad sin precedentes para la innovación, pero también un riesgo significativo: la amplificación y perpetuación de sesgos inherentes en los datos de entrenamiento. Estos sesgos, a menudo sutiles y arraigados en la información histórica, pueden llevar a resultados discriminatorios, injustos o simplemente erróneos, afectando la reputación de la empresa y, crucialmente, a los usuarios finales. Los líderes empresariales, conscientes de que los sesgos de entrenamiento no les afectan directamente a ellos sino a sus clientes y a la sociedad, deben adoptar un enfoque estratégico proactivo para mitigar estos riesgos. Aquí es donde la Nueva Teoría Estratégica (NTE) ofrece un marco valioso.

La Nueva Teoría Estratégica (NTE) como marco para la gobernanza de la IA

23

La Nueva Teoría Estratégica (NTE), propuesta por Rafael Alberto Pérez, redefine la estrategia desde una perspectiva de la comunicación y la complejidad, alejándose de visiones lineales y mecanicistas (CINTE, s.f.). La NTE enfatiza la co-construcción de la realidad, la importancia de los flujos de información y la interconexión de los sistemas. Aplicada a la gobernanza de la IA, la NTE sugiere que la mitigación de sesgos no es un problema técnico aislado, sino un desafío estratégico que requiere:

- Visión holística y sistémica: Reconocer que los LLMs son sistemas complejos que co-evolucionan con sus datos, usuarios y el entorno social. Los sesgos no son fallas aisladas, sino manifestaciones de patrones en los datos y en la interacción humana. La estrategia debe abordar el ecosistema completo, no solo el modelo (CINTE, s.f.).

- Enfoque en la comunicación y el diálogo: La NTE postula que la comunicación es el locus de la estrategia. Para la IA, esto significa establecer canales de comunicación transparentes sobre cómo se entrenan los modelos, qué datos se utilizan y cómo se abordan los sesgos. Implica un diálogo constante con los stakeholders (usuarios, expertos en ética, reguladores) para entender sus preocupaciones y expectativas (Grupo DIRCOM, 2023).

- Principio de recursividad y retroalimentación: La NTE subraya la importancia de los bucles de retroalimentación. En el contexto de los LLMs, esto se traduce en la implementación de mecanismos robustos

para monitorear continuamente el rendimiento del modelo en el mundo real, identificar sesgos emergentes y utilizar esa retroalimentación para refinar los datos de entrenamiento y los algoritmos. Es un proceso iterativo de aprendizaje y adaptación (Herrera, 2021).

- Recuperar al ser humano como centro: La NTE busca recuperar la centralidad del ser humano frente a constructos puramente económicos o tecnológicos. Para los líderes empresariales, esto significa priorizar el impacto humano de los LLMs, asegurando que su desarrollo y despliegue beneficien a todos los usuarios y no perpetúen desigualdades. La ética y la equidad deben ser principios rectores, no consideraciones secundarias (Trujillo, 2019).

Estrategias concretas para la mitigación de sesgos en datos de entrenamiento de LLMs

Basándose en los principios de la NTE y las mejores prácticas en IA ética, los líderes empresariales pueden implementar las siguientes estrategias:

- 1. Curación y diversificación de datos de entrenamiento:

- Auditoría de sesgos en datos existentes: Realizar auditorías exhaustivas de los conjuntos de datos utilizados para entrenar LLMs para identificar y cuantificar sesgos demográficos, culturales, de género, raciales, etc. Esto implica el uso de herramientas automatizadas y revisión humana (Maceiras, 2024).

- Inclusión de datos representativos: Buscar activamente y priorizar la inclusión de datos de diversas fuentes y demografías subrepresentadas. Esto puede implicar la creación de nuevos conjuntos de datos o la ponderación de los existentes para asegurar una representación equitativa (NEWS, 2024).

- Filtrado y limpieza de sesgos explícitos: Eliminar o mitigar el lenguaje ofensivo, discriminatorio o estereotipado presente en los datos. Esto requiere un equilibrio delicado para no eliminar el contexto histórico o social relevante, pero sí evitar la perpetuación de prejuicios dañinos.

- 2. Diseño de algoritmos conscientes de sesgos:

- Algoritmos de equidad: Desarrollar e implementar algoritmos que incorporen métricas de equidad durante el entrenamien-

to, buscando minimizar las diferencias de rendimiento o los resultados sesgados entre diferentes grupos demográficos (The Conversation , 2024).

- Técnicas de Des-sesgo (Debiasing): Aplicar técnicas algorítmicas para reducir los sesgos en los modelos, ya sea en la fase de pre-procesamiento de datos, durante el entrenamiento o en la post-procesamiento de los resultados. Esto puede incluir el uso de adversarios generativos para identificar y corregir sesgos (Dharani et al., 2023).

-3. Transparencia y explicabilidad (XAI):

- Documentación detallada de datos y modelos: Mantener una documentación exhaustiva sobre el origen de los datos, los métodos de curación, las decisiones de diseño del modelo y las métricas de sesgo utilizadas. Esta transparencia es crucial para la rendición de cuentas (Comisión Europea, s.f.).

- Modelos explicables: Desarrollar LLMs que puedan explicar sus decisiones o la base de sus respuestas, permitiendo a los usuarios y desarrolladores entender por qué se generó una salida particular y si hay sesgos subyacentes. Esto es vital para construir confianza y permitir la auditoría externa.

25

- 4. Gobernanza y auditoría externa:

- Comités de ética de IA: Establecer comités internos multidisciplinares (con expertos en ética, sociología, psicología, además de ingenieros) para supervisar el desarrollo y despliegue de LLMs, asegurando que se adhieran a principios éticos y de equidad (Naciones Unidas, 2023).

- Auditorías independientes: Someter los LLMs y sus datos de entrenamiento a auditorías externas e independientes para evaluar la presencia de sesgos y la efectividad de las medidas de mitigación. Esto añade una capa de objetividad y credibilidad.

- Marcos regulatorios y estándares: Participar activamente en el desarrollo de marcos regulatorios y estándares de la industria para la IA ética y la mitigación de sesgos. Las empresas que lideren en este ámbito no solo cumplirán con futuras normativas, sino que también construirán una ventaja competitiva basada en la confianza.

- 5. Educación y concienciación continua:

- Capacitación del personal: Educar a todos los equipos involucrados en el ciclo de vida de los LLMs (desde ingenieros hasta equipos de marketing y ventas) sobre los riesgos de sesgos, su impacto y las mejores prácticas para su mitigación. Fomentar una cultura de responsabilidad y ética en la IA.

- Alfabetización de usuarios finales: Aunque los líderes empresariales no pueden controlar directamente el consumo de información de sus usuarios, pueden contribuir a la alfabetización digital ofreciendo recursos o integrando funcionalidades que ayuden a los usuarios a identificar contenido generado por IA o potencialmente sesgado, como se discutió en la sección sobre ecosistemas inmunes a la desinformación.

Conclusión: hacia ecosistemas digitales resilientes y éticos

26

La desinformación, impulsada por sesgos cognitivos humanos y amplificada por las capacidades de la inteligencia artificial, representa uno de los desafíos más apremiantes de nuestra era digital. La particular vulnerabilidad de los adultos mayores a la información errónea subraya la urgencia de un enfoque proactivo y ético en el desarrollo y despliegue de tecnologías como los LLMs.

Los líderes empresariales tienen una responsabilidad ineludible en este panorama. No solo deben buscar la eficiencia y la innovación a través de la IA, sino también asegurar que estas tecnologías se desarrollen de manera que mitiguen los sesgos y protejan a los usuarios finales. La aplicación de los principios de la Nueva Teoría Estratégica (NTE) ofrece un marco robusto para esta tarea, promoviendo una visión holística, comunicativa y centrada en el ser humano para la gobernanza de la IA.

Al invertir en la curación y diversificación de datos, el diseño de algoritmos conscientes de sesgos, la transparencia, la gobernanza ética y la educación continua, las empresas no solo cumplirán con un imperativo moral, sino que también construirán una ventaja competitiva basada en la confianza y la credibilidad. Solo a través de un compromiso sostenido con la ética y la responsabilidad podremos diseñar ecosistemas digitales verdaderamente resilientes, donde la verdad prevalezca sobre la desinformación y la tecnología sirva al bienestar de toda la sociedad.

Referencias

- AARP Blogs (2022, 25 de octubre). *Aprende a detectar y detener la desinformación en línea*. <https://blog.aarp.org/lucha-por-ti/herramientas-contra-la-desinformacion-en-internet>
- Área Humana. (s.f.). *Bulos y desinformación Claves psicológicas*. <https://www.areahumana.es/bulos-y-psicologia/>
- Blanco-Alfonso I., Solano-Altaba M. y Rodríguez-Luque C. (2024). Actitudes, sesgos cognitivos y sentimientos ante la desinformación a través de redes sociales. *Estudios sobre el Mensaje Periodístico*, 30(3), 467-475. <https://doi.org/10.5209/emp.96800>
- Cantril, H. (1940). *The Invasion from Mars: A Study in the Psychology of Panic*. Princeton University Press.
- CINTE (s.f.). *¿Qué es la Nueva Teoría Estratégica (NTE)?*. <https://nuevateoriaestrategica.digital/la-nueva-teoria-estrategica/>
- CINTE (s.f.). *Fundamentos de la Nueva Teoría Estratégica*. <https://nueva-teoriaestrategica.digital/bienvenida/>
- Comisión Europea (s.f.). *Lucha contra la manipulación de la información*. https://commission.europa.eu/topics/countering-information-manipulation_es
- Daríá, V. (2024, 23 de septiembre). *¿Por qué creemos en noticias falsas? Psicología de la desinformación*. <https://psicologiadesastre.com/psicologia-de-la-desinformacion/>
- Dharani, N., Ludwig, J., & Mullainathan, S. (2023, 18 de enero). Can A.I. Stop Fake News? . *Chicago Booth Review*. <https://www.chicago-booth.edu/review/can-ai-stop-fake-news>
- Hernández, R. (s.f.). *Cómo enseñar a los mayores a identificar la desinformación*. <https://www.connectas.org/labs/como-ensenar-a-los-mayores-a-identificar-la-desinformacion/>
- Herrera, R. (2021). *Nueva Teoría Estratégica: reflexiones y aportes desde la comunicación*. https://www.researchgate.net/publication/359952500_Nueva_Teoria_Estrategica_reflexiones_y_aportes_desde_la_comunicacion
- Grupo DIRCOM. (2023, 1 de diciembre). *Nueva Teoría Estratégica*. <https://grupodircom.com/nueva-teoria-estrategica/>
- Maceiras, S. (2024). *Detección y Clasificación de Fake News mediante Inteligencia Artificial*. http://oa.upm.es/82075/1/TFG_RODRIGO_ESPINOSA_FERNANDEZ_DE_SEVILLA.pdf
- Montiel Torres, M. F. ., Teruel Rodríguez, L., García-Faroldi, L., & Martín-Martín, F. M. (2025). Colectivos vulnerables y desinformación. Análisis de la realidad andaluza. *Doxa Comunicación. Revista Interdisciplinar De Estudios De Comunicación Y Ciencias Sociales*, 41, 369-393. <https://doi.org/10.31921/doxacom.n41a2902>
- Mschools (s.f.) *Sesgos cognitivos: ¿por qué caemos en la desinforma-*

ción?. <https://mschools.com/es/sesgos-cognitivos-por-que-caemos-en-la-desinformacion/>

Naciones Unidas (2023, 6 de noviembre). *Siete principios contra la desinformación y las noticias falsas en las redes sociales*. <https://news.un.org/es/story/2023/11/1525477>

NEWS (2024, 22 de febrero). *AI and the spread of fake news sites: Experts explain how to counteract them*. <https://news.vt.edu/articles/2024/02/AI-generated-fake-news-experts.html>

Parra, S. (2022, 15 de enero). *Adulto mayor y noticias falsas: ¿qué hacer ante la desinformación?* <https://www.javeriana.edu.co/pesquisa/adulto-mayor-y-noticias-falsas-que-hacer-ante-la-desinformacion/>

Pérez Colomé, J. (2025, 19 de enero). *La gente mayor prefiere la información dudosa, no por falta de habilidades digitales sino porque confirma su ideología*. <https://elpais.com/tecnologia/2025-01-20/la-gente-mayor-prefiere-la-informacion-dudosa-no-por-falta-de-habilidades-digitales-sino-porque-confirma-su-ideologia.html>

Radio Ambulante (s.f.). *Cuando la “Guerra de los mundos” llegó a Quito – Fotos y videos*. <https://radioambulante.org/extras/cuando-la-guerra-de-los-mundos-llego-a-quito-fotos-y-videos>

Rivas De La Rosa, Elvis Onasis (2022). *Analfabetismo digital y desinformación en los adultos mayores de la parroquia Santa Rosa*. La Libertad. UPSE. Matriz. Facultad de Ciencias Sociales y de la Salud. 60p. <https://repositorio.upse.edu.ec/handle/46000/8456>

Sádaba, Charo; Salaverría, Ramón; Bringué-Sala, Xavier (2023). “How to teach the elderly to detect disinformation: a training experiment with WhatsApp”. *Profesional de la información*, v. 32, n. 5, e320504. <https://doi.org/10.3145/epi.2023.sep.04>

Scherer, L. D., & Pennycook, G. (2021). ¿Quién es más susceptible a la información errónea en línea sobre la salud?. *Revista Panamericana de Salud Pública*, 45, e53. <https://doi.org/10.26633/RPSP.2021.53>

The Conversation (2020, 20 de agosto). *Los mayores de 50 (y no los jóvenes) son quienes más noticias falsas comparten*. <https://theconversation.com/los-mayores-de-50-y-no-los-jovenes-son-quienes-mas-noticias-falsas-comparten-141649>

The Conversation (2024, 6 de noviembre). *How close are we to an accurate AI fake news detector?*. <https://theconversation.com/how-close-are-we-to-an-accurate-ai-fake-news-detector-242309>

The Decision Lab. (s.f.). *El efecto de la desinformación*. <https://thedecisionlab.com/es/reference-guide/psychology/the-misinformation-effect>

Trujillo, A. (2019, 11 de marzo). *¿Una Nueva Teoría Estratégica?*. <https://cila.comminit.com/content/%C2%BFuna-nueva-teor%C3%A-Da-estrat%C3%A9gica>

Welch, D. (1993). *The Third Reich: Politics and Propaganda*. Routledge.
YLAJ Network. (2019, 1 de julio). *La psicología de la desinformación*. <https://ylai.state.gov/es/la-psicologia-de-la-desinformacion/>

La reputación organizacional en la era digital: un relato de sombras y resiliencia

30

Alejandro Salum

CEO de Tascomm, firma especializada en consultoría para el crecimiento sostenible y la transformación organizacional. Cuenta con una sólida trayectoria en mercadotecnia y desarrollo de negocios a nivel internacional, habiendo ocupado cargos directivos en empresas globales. Destaca su rol como CMO en Atento, donde lideró estrategias de marketing, comunicación y nuevos negocios en múltiples regiones, incluyendo América, Europa y África.

Fue CMO de Microsoft México durante más de cinco años, así como CMO y Director de Nuevos Negocios en Sun Microsystems de México. Ha desempeñado puestos como Presidente y Director General de Grupo Projekta, Director Internacional de Mercadotecnia en Grupo CIE y Director de Mercadotecnia en Burson-Marsteller en Argentina. Actualmente, es Vicepresidente de Estrategia de la Academia Mexicana de la Comunicación y cuenta con formación ejecutiva en prestigiosas instituciones como IPADE, EGADE, Berkeley, Stanford, Kellogg y Harvard.

La era digital ha redefinido el panorama de riesgos para las organizaciones, donde las ciberamenazas y la desinformación, potenciadas por la Inteligencia Artificial Generativa (IAG), representan desafíos sin precedentes para la reputación. Después de más de 30 años de trabajar para empresas transnacionales, velando por la reputación y posicionamiento de productos, servicios, personas y, siempre dirigiendo áreas de comunicación interna y externa, vivo este momento como uno de los más aterradores pero emocionantes de mi carrera como comunicólogo organizacional.

La era digital ha transformado radicalmente el entorno empresarial y social, presentando un doble filo: mientras abre puertas a la conectividad y la eficiencia sin precedentes, también expone a las organizaciones a amenazas digitales cada vez más sofisticadas y devastadoras. Los incidentes y las crisis, lejos de ser eventos aislados, son una constante que “llegan sin avisar” y afectan a todos los sectores, desde los medios y la banca hasta la manufactura y la tecnología. En este contexto, la reputación organizacional se ha convertido en uno de los activos más valiosos y, paradójicamente, uno de los más vulnerables. Como señalara el empresario norteamericano Warren Buffet, “Toma 20 años construir una buena reputación y 5 minutos arruinarla”. Este artículo profundiza en la naturaleza de estas nuevas amenazas digitales, explorando su impacto multidimensional y delineando un camino para que los comunicadores organizacionales desarrollen estrategias robustas que no solo mitiguen el daño, sino que también refuercen la resiliencia reputacional en un entorno de constante evolución.

31

El telón de fondo de nuestra era se dibuja con una paleta de desafíos que se extienden más allá de lo previsible. Desde la escasez que aprieta y el alza de precios que ahoga, hasta los ecos de la corrupción, las cadenas de una legislación restrictiva, la difamación que carcome y, como una marea creciente, la implacable ola de ciberataques. Estos últimos, en particular, han mutado, volviéndose “más complejas, frecuentes e impredecibles”, una hidra de mil cabezas que se manifiesta en ciberdelincuencia, violaciones de datos, fugas de información que se escurren entre los dedos, sabotajes silenciosos, suplantaciones de identidad que confunden, ciberextorsiones que asfixian, infecciones masivas de malware que se propagan como plagas, y asaltos a las infraestructuras en la nube o a las intrincadas cadenas de suministro. Una amenaza constante, un susurro de peligro que nunca cesa.

El eco financiero de estos ciberdelitos resuena con una magnitud asombrosa. Para el año 2025, se proyecta que el costo global de esta delincuencia digital escalará a la vertiginosa cifra de 9.8 trillones

de dólares. Esto se traduce en aproximadamente 301,243 dólares por segundo, con un ataque significativo irrumpiendo cada 4 o 5 segundos (Morgan, 2023). Una cifra que no solo es un número, sino el reflejo de un crecimiento implacable: un 34% anual, y un escalofriante incremento del 400% entre 2020 y 2024 (Morgan, 2023). El Foro Económico Mundial, en su “Global Cybersecurity Outlook 2025”, nos presenta una verdad inquietante: el 60% de las empresas han tropezado con al menos un ciberataque de envergadura en el último año (World Economic Forum, 2025). Entre los ataques más comunes nos encontramos el ransomware, que secuestra datos en el 40% de los ataques; el phishing, que engaña al 30%; y los ataques a la cadena de suministro, que han incrementado en un 20% (World Economic Forum, 2025).

32 Las pérdidas globales por ciberataques en 2024 no fueron menos de 9.4 trillones de dólares (Gutiérrez, 2025). Y, en un eco aún más desolador, el 60% de las Pequeñas y Medianas Empresas (PYMEs) españolas, por citar un ejemplo, que cayeron víctimas de ciberataques en 2024, cerraron sus puertas en los seis meses siguientes al incidente, un testimonio crudo de la fragilidad de estas organizaciones ante la embestida cibernética (Gutiérrez, 2025). La proliferación de dispositivos del Internet de las Cosas (IoT), ha extendido aún más el campo de batalla; un asombroso 98% del tráfico de estos dispositivos permanece sin cifrar, una puerta abierta de par en par, y los ataques de malware en IoT se dispararon un 400% en 2023 (Gutiérrez, 2025). Se estima que unos 17.000 millones de dispositivos IoT pueblan nuestro mundo, y una porción considerable de ellos podría convertirse en un talón de Aquiles, vulnerable a la explotación a medida que su soporte y actualizaciones de seguridad se desvanecen en el tiempo (Gutiérrez, 2025).

La irrupción y el vertiginoso ascenso de la Inteligencia Artificial Generativa (IAG) han añadido una capa de complejidad, un velo casi místico, a las amenazas digitales. Si bien la IA promete un sinfín de beneficios, como un genio liberado de su lámpara, también ha abierto la caja de Pandora a riesgos reputacionales nuevos y profundos para las organizaciones. He observado con preocupación que en las empresas la capacitación en relación con el uso y adopción de estas nuevas tecnologías es extremadamente básica, lenta y/o inexistente. ChatGPT (la IA de preferencia de millones) se ha convertido en el amigo, confidente, chef, profesor, asesor, aliado matemático, redactor, creativo, etc., de profesionistas quienes continuamente alimentan al ecosistema con información personal, familiar y empresarial abriéndose al mundo con capa ancha e ignorando los riesgos que esto implica, riesgos que entre otros destacan:

- La desinformación y la sombra de los *Deepfakes*: La habilidad de la IAG para tejer contenido sintético con un realismo asombroso ha desatado una era donde las narrativas falsas o engañosas pueden propagarse con una credibilidad alarmante. Los ciberdelincuentes, con esta nueva herramienta en sus manos, están orquestando ataques más precisos y sofisticados, puliendo sus técnicas de suplantación de identidad y forjando mensajes tan convincentes que la efectividad de los ataques de phishing y otras amenazas se dispara (Gutiérrez, 2025).

- La violación de la privacidad y el uso indebido de datos: La IA, en manos equivocadas, puede convertirse en una llave maestra para comprometer datos personales y empresariales, desatando una cascada de consecuencias legales y una erosión irreparable de la confianza.

- *Phishing*, *Vishing* y *Smishing*: La evolución del engaño: Las técnicas de ingeniería social se visten de gala con la IA, volviéndose más astutas y difíciles de detectar, personalizando ataques a una escala masiva que antes era impensable.

33

- La trampa de la dependencia excesiva: Una confianza ciega y sin regulación en los sistemas de IA puede conducir a vulnerabilidades operativas y a dilemas éticos de proporciones colosales.

- *Fake News* y Contenido Inapropiado: La plaga de la mentira: La difusión descontrolada de información falsa o dañina puede escalar con la velocidad de un incendio forestal, calcinando la percepción pública de una organización en cuestión de instantes.

La ciberdelincuencia, ahora con el respaldo de la IAG, la automatización que no descansa, la hiperpersonalización que nos conoce mejor que nosotros mismos, el cómputo cuántico que desafía la comprensión y la ingeniería social que se desliza por las redes sociales, se perfila como la “más peligrosa amenaza cibernética y reputacional de cara al futuro” (Salum, 2025). Una sombra que se alarga, prometiendo desafíos aún mayores en el horizonte.

Los ciberdelitos no son meros ladrones de arcas; su estela va más allá de las pérdidas financieras directas, pues “dañan directamente a las empresas y afectan de forma crítica su reputación” (Salum, 2025). Las consecuencias, como las ondas en un estanque, son múltiples y severas, extendiéndose en círculos concéntricos:

- La lealtad y la confianza de los clientes: La pérdida de datos, ese tesoro digital, o una respuesta titubeante y deficiente ante un incidente,

erosiona la confianza del cliente como la arena bajo el viento, dejándolo expuesto y vulnerable.

- Desinformación, malinformación o información errónea: Estos fenómenos, como espejismos en el desierto, pueden distorsionar la narrativa, tejiendo un velo de confusión que exagera la crisis hasta límites insospechados.

- Consecuencias legales y regulatorias: Tras una brecha de seguridad, las multas, las demandas y el escrutinio regulatorio se ciernen como nubes de tormenta, comunes y a menudo implacables.

- Disrupción operativa: Los ataques, cual parásitos digitales, pueden paralizar las operaciones, afectando la productividad y deteniendo el pulso vital de la organización.

- Daño reputacional: Una herida profunda y duradera en la imagen pública de la organización, una mancha que el tiempo tarda en borrar.

- Pérdidas financieras: Más allá de los costos directos, la factura incluye oportunidades perdidas, sueños truncados y un aumento inexorable en los costos de capital.

Los costos financieros, como un grito de alarma, son escalofriantes. El costo promedio total de una brecha de datos se elevó a 4.45 millones de dólares en 2024, con variaciones que dibujan un mapa de vulnerabilidad por sector (salud: \$10.1M; financiero: \$6.2M; energía: \$5.8M) (Salum, 2025). Estos costos se desglosan en:

- Costos Directos: La respuesta técnica, un promedio de \$1.2M; los pagos de rescate, una media de \$925,000; la notificación a los afectados, \$740,000; y los servicios legales, un promedio de \$1.5M (Salum, 2025).

- Costos indirectos: La pérdida de productividad, que se extiende por varios días las oportunidades perdidas, que duplican los costos directos; el aumento en las primas de seguros, un 35%; y un incremento en el costo de capital de 150 puntos base (Salum, 2025).

El impacto reputacional, un eco aún más devastador, es a menudo un fantasma difícil de atrapar y cuantificar a corto plazo:

- Tasa de abandono de clientes: Un promedio del 3.9% de los clientes, como aves migratorias, pueden abandonar la empresa (Salum,

2025).

- Aumento en el costo de adquisición: El esfuerzo por atraer nuevas almas a la causa puede encarecerse un 23% (Salum, 2025).

- Tiempo de Recuperación: La senda hacia la recuperación de la reputación puede ser larga y tortuosa, extendiéndose de 12 a 24 meses (Salum, 2025).

- Caída en valor de mercado: Una reducción inmediata del 7.5% en el valor de mercado, un golpe que resuena en los pasillos de la bolsa (Salum, 2025).

El sombrío relato del gigante de casinos MGM ilustra con cruda precisión estas cifras, con un costo estimado de 100 millones de dólares, la exposición de 37 millones de datos personales y 40 millones de dólares en juicios (Salum, 2025).

Los daños reputacionales provocados por ciberataques pueden precipitar una caída promedio del 27% en el valor para los accionistas (Modini, 2025). En la dinámica América Latina, el 25% de las empresas que sufrieron un ciberataque en los últimos dos años confesaron un impacto directo en su reputación (Kaspersky, 2025). Este daño se manifiesta en la desconfianza del mercado, la huida de clientes (16%), el retiro de inversionistas (10%) y cicatrices económicas como la cancelación de pedidos o la caída en el valor de las acciones (10%) (Kaspersky, 2025).

35

Las redes sociales, en este escenario de drama digital, se revelan como “amigas y enemigas” a la vez. Las 4S’s como siempre explico: Su velocidad (Speed), su naturaleza siempre despierta (Sleepless), su capacidad de propagación (Spread) y su escala (Scale) las transforman en un amplificador perpetuo de cualquier crisis informática, donde la información (o la desinformación) puede volverse viral en cuestión de minutos (Salum, 2025).

Ante este “caos digital” que se cierne, la pregunta fundamental para los comunicadores corporativos resuena con urgencia: ¿Es posible domar la cobertura mediática, contener la replicación en las redes sociales, desvelar la distorsión de los hechos y, en última instancia, mitigar el impacto reputacional? La respuesta, como un eco de esperanza, es afirmativa: “Si se puede influir en la conversación, pero no domar la cobertura” (Salum, 2025). Para ello, la acción no es una opción, sino un imperativo; la inacción, el mayor de los riesgos.

La comunicación estratégica debe cimentarse en un enfoque proactivo y metódico, como un estrategia que traza su plan de batalla:

- Nunca permitas que otro hable por ti: Asumir el timón de la narrativa es, en este mar revuelto, una necesidad ineludible.

- Elige tus canales con sabiduría: Seleccionar las plataformas adecuadas para cada mensaje es como elegir el arma correcta para cada combate.

- Detecta con agudeza: Identificar con rapidez “qué, quién, cómo, cuándo, dónde” se gestó el incidente, a través de una vigilancia constante del análisis de canales y de la conversación, es adelantarse a la viralización, es ver la tormenta antes de que llegue.

- Comunicar el pulso: Definir el tono y la manera, asegurar la transparencia que disipa las sombras. Nunca mentir y nunca ocultarse. Establecer la frecuencia y repetición de los mensajes, es dar forma al eco de la verdad. Desarrollar una precisa plataforma de mensajes.

- Responde con prontitud y coherencia: Actuar de manera oportuna y con una voz unificada es la clave para restaurar la calma.

- Considera cada paso: Mantener el canal abierto, monitorear, escalear y elevar a los diferentes niveles de la organización, así como de forma externa según sea necesario, involucrando a todas las audiencias y jugadores relevantes, es tejer una red de seguridad.

- Controla la narrativa: Gestionar la información y la interacción para influir en la percepción es el arte de moldear la realidad en la mente de la audiencia.

La encrucijada entre los ciberincidentes y la desinformación no se resuelve con la improvisación, sino con la preparación. Ésta, como un entrenamiento riguroso, implica:

- Practicar, practicar, practicar (Simulacros): Realizar ejercicios regulares, como ensayos de una obra crucial, para poner a prueba los protocolos de respuesta comunicacional y afinar cada movimiento. Así como detectar y capacitar voceros internos expertos y articulados que transmitan con eficacia y precisión los mensajes institucionales.

- Comprender, educar y capacitar de forma constante y continua: Asegurar que cada miembro del equipo, especialmente los comunica-

dores, estén al tanto de las últimas amenazas y de las mejores prácticas, es armarlos con el conocimiento necesario.

- Protocolos de comunicación interdepartamental para la prevención, manejo, mitigación y recuperación de crisis reputacional: Establecer guías claras para cada fase de la crisis es como tener un mapa detallado en medio de la tormenta.

- Blindaje tecnológico: Invertir en soluciones de ciberseguridad robustas, como murallas inexpugnables, que sirvan como primera línea de defensa.

- Monitoreo y análisis permanente del espectro comunicacional: Mantener una vigilancia constante de lo que se dice sobre la organización en todos los canales, es escuchar el pulso del mundo digital.

- Campañas internas y sensibilización general, gerencial y directiva: Fomentar una cultura de ciberseguridad y conciencia reputacional en todos los niveles de la organización, es sembrar la semilla de la resiliencia en cada corazón.

37

En este complejo tapiz digital, donde las amenazas mutan y la información fluye como un río desbordado, la forma en que concebimos la estrategia debe evolucionar. No podemos aferrarnos a mapas antiguos diseñados para paisajes estáticos. Necesitamos una visión que abrace la fluidez, la interconexión y la esencia humana de cada interacción. Es una invitación a trascender los enfoques tradicionales, a menudo rígidos y lineales, y a adoptar una perspectiva que valora la adaptabilidad, la proactividad y la colaboración en un mundo de incertidumbre creciente.

Desde esta nueva mirada, la comunicación estratégica frente a las amenazas digitales deja de ser una mera táctica para convertirse en el corazón mismo de la estrategia organizacional. Esto se manifiesta en varios pilares:

- La visión del ecosistema interconectado: Debemos reconocer que los ciberincidentes y los daños a la reputación no son eventos aislados, sino hilos entrelazados en una compleja red. La respuesta exige una sinfonía coordinada entre la tecnología, las operaciones, el ámbito legal y, por supuesto, la comunicación. Es entender la organización como un ser vivo, en constante diálogo con su entorno digital, donde la estrategia no es un plan preestablecido, sino una melodía que emerge de procesos fluidos, complejos y, a veces, caóticos.

- La danza de la anticipación: En este escenario, la inteligencia y la capacidad de adelantarse y contar con un análisis de riesgos son cruciales. El monitoreo constante del vasto espectro comunicacional y la habilidad de “adelantarse a la viralización” se convierten en faros que nos guían. Nos permiten percibir las señales más tenues y prepararnos antes de que una amenaza se convierta en una tormenta. Es operar en un entorno de turbulencia constante, donde la anticipación no es un lujo, sino una necesidad vital.

- Agilidad y adaptabilidad: El Arte de Fluir: En un mundo donde las amenazas son “más complejas, frecuentes e impredecibles” (Salum, 2025), la rigidez estratégica es un ancla que nos condena. Necesitamos la flexibilidad en nuestros protocolos de comunicación y la destreza para ajustar nuestras respuestas en tiempo real, tejiendo la estrategia con el hilo del análisis continuo de la situación. Es un proceso dinámico y adaptable, como un río que encuentra su propio cauce.

38 - La orquesta de voces: gestión de “stakeholders”: La reputación no es una fortaleza solitaria, sino un tapiz tejido con las interacciones de múltiples audiencias. La comunicación estratégica debe centrarse en cada parte interesada, modulando la transparencia, la frecuencia y el tono para resonar con las necesidades y percepciones de cada grupo: desde los clientes que confían hasta los reguladores que observan y los empleados que construyen. Es recuperar lo humano, viendo a los individuos no como meros constructos económicos, sino como seres relacionales, con sus historias y sus voces.

- La verdad como escudo inquebrantable: La batalla contra la desinformación y los deepfakes exige una narrativa estratégica tan sólida como una roca, donde “la verdad es el escudo”. La coherencia y la credibilidad de nuestra narrativa oficial son el cimiento para preservar la confianza y disipar la proliferación de información errónea. Es ir más allá del mero beneficio, hacia la sostenibilidad que perdura y la responsabilidad ética que nos guía.

- El alma de la organización: cultura de resiliencia y gobernanza: Debemos tejer la ciberseguridad y la gestión reputacional en el entramado de la gobernanza corporativa, cultivando una “sensibilización general, gerencial y directiva” (Salum, 2025). Esto implica la creación de un “Comité de Seguridad” que no solo mire a la tecnología, sino que también abraza la comunicación estratégica ante los riesgos emergentes. La organización se transforma en un nódulo de innovación y significación, trascendiendo la mera función de unidad de producción.

En esencia, esta nueva perspectiva estratégica nos revela que la defensa reputacional no es una tarea solitaria o una reacción tardía de los comunicadores, sino una responsabilidad compartida que impregna cada fibra de la organización, impulsada por una visión que mira al futuro y una capacidad de adaptación constante frente a la disrupción digital.

Los desafíos que se ciernen sobre los comunicadores organizacionales en la era digital son inmensos. La ciberdelincuencia y la Inteligencia Artificial Generativa convergen, tejiendo un tapiz de amenazas sin precedentes. El eco de un ciberincidente resuena mucho más allá de las frías pérdidas financieras directas; erosiona la confianza, desgarrar la lealtad de los clientes y, en última instancia, disminuye el valor de mercado de una empresa, como una vela que se consume.

Sin embargo, frente a este panorama que podría parecer desolador, la inacción es el abismo más profundo. Los comunicadores corporativos poseemos la capacidad, la chispa, para influir en la conversación, mitigar el impacto negativo y proteger ese tesoro llamado reputación. Esto exige un compromiso inquebrantable con una defensa reputacional activa, donde la preparación rigurosa, el monitoreo constante, la agilidad en la respuesta y una comunicación transparente sean los pilares inamovibles de la estrategia. Como hemos mencionado anteriormente, la adopción de principios estratégicos que abrazan la complejidad y la adaptabilidad es fundamental para navegar las sombras que acechan y asegurar la resiliencia en un campo de batalla donde la reputación es el objetivo final, el premio más codiciado.

39

En la encrucijada entre los ciberincidentes que golpean y la desinformación que confunde, nuestro compromiso debe ser forjar una defensa reputacional activa “con la verdad como escudo y la colaboración como arma”. Solo así, las organizaciones no solo lograrán sobrevivir, sino que florecerán, se alzarán victoriosas en este desafiante y fascinante entorno digital.

Referencias

- Gutiérrez, N. (2025). Estadísticas de ciberseguridad que debes conocer. *Prey Project*. <https://preyproject.com/es/blog/estadisticas-seguridad-informatica>
- Kaspersky (2025, 2 de julio). El 25 % de las empresas en América Latina ve afectada su reputación tras un ciberataque, revela estudio. <https://latam.kaspersky.com/about/press-releases/el-25-de-las-empre->

[sas-en-america-latina-ve-afectada-su-reputacion-tras-un-ciberataque-revela-estudio?srsId=AfmBOoqVMYuFCSEYOpVIJxQoWlt-NYCNIQWbmCdngp04lUYqjduul7B0W](https://www.innovaciondigital360.com/cyber-security/las-cifras-que-muestran-como-los-danos-reputacionales-por-ciberataques-golpean-el-valor-de-las-empresas-en-todo-el-mundo/)

Modini, T. (2025, 25 de agosto). Las cifras que muestran cómo los daños reputacionales por ciberataques golpean el valor de las empresas en todo el mundo. *Innovación Digital 360*. <https://www.innovaciondigital360.com/cyber-security/las-cifras-que-muestran-como-los-danos-reputacionales-por-ciberataques-golpean-el-valor-de-las-empresas-en-todo-el-mundo/>

Morgan, S. (2023, 25 de octubre). Cybercrime to cost the world \$9.5 trillion USD annually in 2024. *Cybersecurity Ventures*. <https://cybersecurityventures.com/cybercrime-to-cost-the-world-9-trillion-annually-in-2024/>

Salum, A. (2025, 3 de septiembre). Reputación bajo ataque: Comunicación estratégica ante las nuevas amenazas digitales. *Presentación en el evento FUTURIBLE 2025 CiNTE*. [Diapositivas 1-13].

World Economic Forum (2025). *Global Cybersecurity Outlook 2025*. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf

MESA

Gestión de la desinformación y la credibilidad en la política y la gobernanza

2

41

Moderadora

María Eugenia Garcés

Licenciada en Comunicación Social, Magíster en Comunicación Empresarial. Docente en la Facultad de Comunicación en la Universidad Central del Ecuador. Directora de Comunicación Organizacional y Estratégica en organismos públicos. Consultora de Comunicación en proyectos auspiciados por Organismos Internacionales: PNUD, UNESCO, UNFPA, BID, Banco Mundial, en temas de comunicación intercultural, comunicación ambiental, comunicación y derechos humanos. Asesora de comunicación en organizaciones sociales, especialmente indígenas y ambientales. megarces@uce.edu.ec

La vorágine informativa del siglo XXI y especialmente de su segunda década, nos está enfrentando a una visión caótica, desordenada y conflictiva del mundo. La pandemia del COVID 19 marcó un hito, luego del cual todo se convirtió en un caos, al ser aprovechado para reorientaciones políticas y económicas que subyacían, provocando profundas tensiones.

Mientras la tecnología, especialmente la de información y comunicación avanza vertiginosamente, la política, la economía, las relaciones sociales están siendo aplicadas para un reacomodo de hegemonías, que han desencadenado procesos inenarrables que no pensamos que se fueran a repetir, como el genocidio en Gaza, la beligerancia entre potencias, la posibilidad siempre existente de una nueva guerra mundial, con el agravante del uso de armas nucleares; la creciente pobreza y lacerante desigualdad social, sobre todo en los países subdesarrollados, sometidos a políticas que atentan contra los derechos y la dignidad humana; a lo que se suma la insostenible explotación de recursos naturales para garantizar la imparable reproducción del capital, que ha puesto en riesgo la existencia de la humanidad misma.

En medio de ello una necesidad social imperiosa de descubrir mecanismos que permitan espacios de encuentro, donde se puedan establecer consensos para beneficio de las poblaciones, más allá de la disputa de quienes detentan los poderes políticos y económicos y los usan para su exclusivo beneficio, imponiendo políticas concentradoras y que impiden la redistribución de la riqueza para que la población pueda desarrollarse con dignidad y respeto.

La discusión sobre la democracia, la participación igualitaria y los derechos, han sido desaparecidos, cuando no denostados por burbujas de desinformación, estratégicamente formuladas y que tienen como objetivo disociar, enemistar, desencontrar. Cada vez sentimos más lejano el consenso y la gobernabilidad, como mecanismo de relación, de búsqueda de acuerdos políticos entre los diversos, que garanticen el ejercicio de gobiernos aceptados socialmente, en los que prime el lenguaje de la negociación, con la finalidad de construir articulaciones.

En estas circunstancias recuperar y ampliar un sistema político que medie entre la sociedad y el Estado, con capacidad de articular un proyecto participativo para el presente y futuro, que establezca canales de interacciones, relaciones y redes entre los distintos sectores (gubernamental, sector público, privado y la sociedad en general), mecanismos participativos de toma de decisiones, negociación y relaciones equilibradas de poder, constituye en la sociedad actual una obligación ética, para

alejar los ejercicios de poder autoritarios.

Sumado a ello, la gobernanza como requisito indispensable para el ejercicio del derecho a la participación ciudadana de la llamada sociedad civil, las instituciones y sectores organizados de distinta naturaleza, que se da a través de mecanismos de autorregulación para establecer acuerdos entre diversos, permitiendo la representación política de todos los interactuantes en la toma de decisiones que les afectan directamente, parece venida a menos. Sociedades que no logran reconocerse ni política, ni socialmente no pueden establecer puntos de encuentro, perdiendo la posibilidad real de influir en las decisiones que adoptan los gobiernos, en su nombre.

Lo paradójico es que todo lo aquí señalado tiene como eje la disociación, la ruptura de las interacciones, generadas en gran medida por una estampida de desinformación en la era de la información. Esta desinformación que constituye la difusión falsa o engañosa, emitida de manera consiente y deliberada planeada estratégicamente para manipular a las audiencias y orientar tendencias e influenciar en la opinión pública. Su acción política se orienta a provocar la desconfianza en la institucionalidad del Estado y los medios de comunicación social, que tienen la capacidad de amplificar voces diversas, dando como resultado la legitimación de discursos y prácticas autoritarias, lo que pone en riesgo el funcionamiento de la democracia.

43

La utilización de la desinformación en el mercado del marketing político promueve ecosistemas mediáticos polarizados. La construcción de enemigos y opositores, no desde dimensiones ideológicas sino afectivas, cierran los caminos de la deliberación pública racional y la posibilidad de llegar a negociaciones y más bien generan hostilidades, al parecer insalvables, dando paso a gobiernos con tintes populistas y autoritarios.

Para profundizar en este tema, en el marco de FUTURIBLE 2025, tenemos la participación de Gabriel Salazar Navarro, abogado de la Universidad de La Sabana.

El desafío de gobernar en medio de narrativas virales

44

Gabriel Salazar Navarro

Especialista en Derecho de la Competencia y del Libre Comercio por la Pontificia Universidad Javeriana, y magíster cum laude en Comunicación de la Universidad Sergio Arboleda, donde su tesis de maestría —“Protesta Social en Tiempos de Pos-verdad y Redes Sociales”— recibió distinción meritoria por parte del jurado académico.

Cuenta con más de 16 años de experiencia profesional —13 de ellos dedicados a la gestión de bienes públicos— y ha liderado proyectos institucionales que integran Derecho, Administración, Tecnología y Comunicación. Ha participado en la formulación de políticas públicas, el diseño normativo y el desarrollo de soluciones innovadoras para el sector justicia.

Es autodidacta, editor de libros como *Tsunami 1979* y autor de *La noche en que Bogotá se tragó sus propios miedos*. Promotor del uso responsable de tecnologías, actualmente administra el sitio web www.gabrielsalazarnavarro.com, donde comparte contenido formativo y editorial sobre sociedad, tecnología, comunicación y administración pública.

1. De Tesich a Trump, anatomía histórica y conceptual de la posverdad

La palabra posverdad fue elegida como término del año en 2016 por el Diccionario Oxford y se incorporó a la versión digital del Diccionario de la Real Academia Española en el año 2017. Con este concepto se ha designado la época que atravesamos, marcada por el hecho de que *“las personas responden más a sentimientos y creencias que a hechos”* (Oxford Languages, 2016). Sin embargo, hoy, en pleno 2025 y ocho años después de su consagración, todavía son muchos los que desconocen el verdadero alcance de esta noción o no han reparado en lo que implica vivir en tiempos de posverdad.

De entrada conviene subrayar que no se trata de un simple recurso retórico y que los términos posverdad y fake news alcanzaron su mayor notoriedad en un escenario político de alta tensión, esto es, durante las elecciones presidenciales estadounidenses del año 2016, por lo que no es casual que gran parte de su popularidad se asocie a Donald Trump; De hecho, el uso sistemático de la etiqueta “fake news” en sus discursos, trinos e anteverciones transformaron el campo discursivo y la relación con la prensa tradicional en Estados Unidos (Ross & Rivers, 2017).

45

No obstante, la palabra posverdad no nació ni con Donald Trump ni en el año 2016, su primera aparición documentada en el sentido actual, según el rastreo del Diccionario Oxford, se atribuye al guionista y dramaturgo Steve Tesich, quien la empleó en 1992 en su *artículo “The Watergate Syndrome: A Government of Lies”*, publicado en la revista The Nation.

Con esa invención léxica, que con el tiempo se convertiría en neologismo, Tesich buscó describir lo que diagnosticó como “The Watergate Syndrome”, un padecimiento colectivo de la ciudadanía estadounidense que, ante crisis políticas o revelaciones mediáticas impactantes, prefería evadir la verdad antes que enfrentarse a ella, pues los hechos verificados por la prensa les resultaban contrarios a sus dogmas, increíbles o simplemente insoportables de aceptar (Tesich, 1992).

Vale la pena destacar que los tres episodios de alto impacto y cuestionable gestión gubernamental a los que Tesich aludió, para mostrar cómo se configura una negación colectiva, fueron el Watergate (Bernstein & Woodward, 1974), el Irán-Contra (Kornbluh, 1993) y la Guerra del Golfo Pérsico (Kellner, 1992), hechos que constituyen hitos políticos de amplia exposición mediática de la historia estadounidense, y que conviene examinarlos porque son propios de una época previa

a la existencia de las redes sociales y a los recursos que posibilitan su propagación en el ecosistema digital contemporáneo.

Fue a partir de esos antecedentes, que Tesich escribió casi que proféticamente lo siguiente: «*Nos estamos convirtiendo rápidamente en prototipos de un pueblo del que los monstruos totalitarios solo podrían babear en sus sueños. Todos los dictadores hasta ahora han tenido que trabajar duro para suprimir la verdad. Nosotros, con nuestras acciones, estamos diciendo que esto ya no es necesario, que hemos adquirido un mecanismo espiritual que puede despojar a la verdad de cualquier significado. **De una manera muy fundamental, nosotros, como personas libres, hemos decidido libremente que queremos vivir en un mundo posterior a la verdad***». (Tesich, S.1992) (subraya y negrilla fuera del texto original)

Esos tres acontecimientos políticos, que revisaremos brevemente a continuación, resultan en clave comunicacional coherentes con lo que autores de la Nueva Teoría Estratégica y de la Comunicación Estratégica Enactiva vienen planteando desde hace algún tiempo, que la estrategia se refunda “desde la comunicación y lo relacional” (Pérez, 2014) y que la comunicación es un encuentro intersubjetivo en el que emociones y creencias pueden pesar más que los datos en la construcción de legitimidad (Massoni, 2019) en el plano ético, además, esos mismos casos marcados por crisis de confianza y encubrimiento, remiten a lo que Joan Costa planteó en su columna *¿Qué es la verdad?*, publicada en *Experimenta*, donde advierte que la sociedad de consumo y del espectáculo convierte todo en apariencia y simulacro, incluso la justicia y la política, dejando la verdad como un artificio frágil, sostenido solo mientras resiste el paso del tiempo o la manipulación de la opinión pública (Costa, 2021).

1.0.1. El caso Watergate (1972)

Tomemos como marco de referencia el libro *All the President's Men* (1974), de Carl Bernstein y Bob Woodward, relato detallado de quienes realizaron la investigación periodística que destapó el escándalo. Watergate es un eje clave para Tesich, de hecho, con esa palabra titula su reflexión y sintetiza la deriva cultural y política de Estados Unidos tras el episodio, que deja a una democracia golpeada por la caída de su presidente y que él marca, como el inicio de la configurar de un patrón ciudadano frente a los escándalos de gobierno.

El caso estalló la madrugada del 17 de junio de 1972 con el arresto de cinco intrusos en la sede del Comité Nacional Demócrata, en el complejo Watergate (Washington, D. C.). Lo que parecía un robo torpe

resultó ser la punta de una operación política clandestina vinculada a la campaña de reelección de Richard Nixon (el “Committee to Re-elect the President”).

Desde la redacción de The Washington Post, Woodward y Bernstein reconstruyeron el caso con verificaciones documentales, entrevistas cruzadas y la guía de una fuente de alto nivel, “Garganta Profunda”, que proporcionaba pistas verificables (Bernstein & Woodward, 1974); El corazón técnico de su indagación fue el “*follow the money*”, que consistía en seguir los cheques, las cuentas y los pagos en cadena para vincular operativos con la estructura formal de la campaña Nixon, y a ello se iban sumando evidencias de presiones a testigos, obstrucción a la justicia y encubrimiento desde altas esferas del Gobierno.

El punto de no retorno llegó en julio de 1973, cuando se reveló un sistema secreto de grabación en el Despacho Oval, y fueron aquellas cintas las que transformaron un “robo” en una crisis constitucional (Bernstein & Woodward, 1974). La presión mediática, política y judicial no se hizo esperar, y fue tal que Nixon renunció el 9 de agosto de 1974 para evitar un “impeachment”. Semanas después, el presidente Gerald Ford le concedió un indulto; varias figuras de su círculo (H. R. Haldeman, John Ehrlichman, John Mitchell) terminaron condenadas.

Este episodio histórico fundamenta la crítica de Tesich, quien, al vivir ese contexto, observaba cómo la verdad era inmolada por narrativas oficiales que buscaban evadir la realidad desde el propio Gobierno, mientras la sociedad permanecía pasmada ante la gravedad de la información revelada por los medios. Esa relación tóxica entre poder y ciudadanía erosionaba las instituciones democráticas hasta reducirlas a meras fachadas, inertes (en principio) frente a semejante crisis de confianza y encubrimiento.

En contraste, el periodismo sacó la verdad a flote, dejó constancia social y configuró una nueva forma de hacer investigación periodística, demostrando que un trabajo riguroso podía derribar a un presidente, incluso frente a toda la fuerza narrativa del Ejecutivo. Watergate estableció así un precedente, graduó a la prensa como cuarto poder dentro del equilibrio constitucional (Robinson, 1974) consolidándola como fiscalizadora del poder y contrapeso legítimo del sistema, y a partir de entonces, la noción de “escándalo político” se consolidó en la opinión pública como arma de control social (Thomas, 1978)

1.0.2. Irán-Contra (1985)

Para el examen de este contexto, usaremos como marco de referencia el texto *"The Iran-Contra Scandal: The Declassified History"* de Kornbluh, P., & Byrne, M. de 1993; El llamado Irán-Contra o Irangate en guiño a lo sucedido años atrás en el Watergate, reveló cómo la manipulación encubierta de fondos, la triangulación de armas y el ocultamiento sistemático de información al Congreso y a la ciudadanía comprometieron la credibilidad democrática de los Estados Unidos, bajo la administración de Ronald Reagan.

La cuestión es que, pese a dos restricciones expresas del Congreso estadounidense (a saber, la prohibición de financiar a los Contras en Nicaragua y el embargo de armas a Irán), altos funcionarios de la administración Reagan diseñaron un plan para vender armas en secreto a Irán, con la coartada de liberar rehenes en Líbano, y destinar esas ganancias al financiamiento encubierto de los Contras en Nicaragua. Todo ello se llevó con el objetivo de sostener una política exterior afín a los intereses de la Casa Blanca y, al mismo tiempo, sortear las limitaciones impuestas por el Legislativo.

El plan se mantuvo oculto durante meses, hasta que en 1986 la revista libanesa *Ash-Shiraa* reveló las ventas de armas, y de ese hilo tiraron *The New York Times* y *The Washington Post*, que días después ampliaron la información y expusieron filtraciones que dejaban al descubierto una red secreta organizada por miembros del Consejo de Seguridad Nacional de Estados Unidos. Entre ellos se encontraba el coronel Oliver North, pieza clave en esta triangulación y en la desobediencia deliberada a las normas impuestas por el Legislativo.

Una vez más la ciudadanía quedaba atónita al descubrir que su propio gobierno había llevado a cabo operaciones encubiertas en abierta contradicción con las prohibiciones expresas del Congreso, pero la verdad, impulsada por la investigación periodística y confirmada en las audiencias televisadas, demostraba que la narrativa oficial podía quebrarse mediante pruebas, documentos y testimonios expuestos ante la opinión pública.

Este episodio se convirtió en una nueva crisis de confianza democrática, y una vez más la comunicación se revelaba como ese encuentro intersubjetivo en el que emociones y creencias pesaban tanto como los datos en la construcción de legitimidad, no obstante, el caso Irán-Contra dejó en claro que la transparencia informativa es una condición indispensable para sostener la confianza ciudadana en las

instituciones y preservar la credibilidad de la democracia (Kornbluh, 1993).

1.0.3. La Guerra del Golfo Pérsico (1991)

Fue en 1991 cuando CNN consolidó su reputación mundial, pues fue esta cadena la que transmitió en vivo y de manera exclusiva el inicio de la Guerra del Golfo Pérsico. Este conflicto representó un punto de inflexión en la historia de la comunicación contemporánea, pues por primera vez una guerra se transmitía en directo a escala planetaria gracias al desarrollo de la televisión satelital.

La cobertura en tiempo real llevó las imágenes del conflicto a millones de hogares alrededor del mundo y generó la sensación de que la guerra podía experimentarse minuto a minuto sin salir de casa, innovación tecnológica que no solo transformó la manera en que se narraban los conflictos armados, sino que también alteró profundamente la relación entre política, periodismo y opinión pública (Kellner, 1993).

49

Esta cobertura satelital casi ininterrumpida, estuvo marcada por un fuerte control informativo, ya que los periodistas o corresponsales de guerra dependían en gran medida de los comunicados oficiales y de las imágenes que les proporcionaba o les permitía hacer las fuerzas militares de la coalición liderada por los Estados Unidos (Taylor, 1992).

En consecuencia, la percepción del conflicto se configuró principalmente a partir de la narrativa gubernamental, en la que predominaba una visión tecnificada y espectacularizada de la guerra moderna, que, de hecho, al revisar en YouTube las transmisiones en vivo de la época, se observa cómo los bombardeos aparecían en pantalla como destellos verdes en la noche captados por cámaras de visión nocturna, mientras aviones, blindados y artillería moderna desfilaban ante la audiencia global.

Todo este espectáculo producía la sensación de estar frente a escenas cuidadosamente diseñadas, más cercanas a un videojuego o a una puesta en escena cinematográfica que a la crudeza de un conflicto bélico. De cualquier manera, lo que transmitían esas imágenes era una exhibición controlada del poderío militar estadounidense, una representación mediática que ocultaba la crudeza del conflicto y que instaló la noción de la guerra como un “espectáculo televisivo” (Kellner, 1993).

Un apunte al margen de este contexto, que quizás podría usarse como recurso más adelante en este examen sobre democracia, gober-

nanza, comunicación y posverdad, es que la cobertura de la Guerra del Golfo contribuyó a consolidar lo que más tarde se conocería como el “CNN effect”, es decir, la idea de que los medios globales (en particular CNN) poseen la capacidad de incidir en la agenda política y en las percepciones ciudadanas durante un conflicto internacional (Robinson, 2002).

Cierro de esta manera el contexto de la Guerra del Golfo Pérsico de 1991, utilizado por Tesich, porque es importante advertir que al igual que en el pasado, las guerras en la actualidad no solo se libran en el terreno militar, sino también en las pantallas de los espectadores, donde la narrativa predominante encuentra eco inmediato y global. Ejemplos contemporáneos (ya en plena era de las redes sociales) son los conflictos entre Rusia y Ucrania (Kreab, 2022) o Israel y Hamas (Ruiz-Baena, 2023). Desinformación sobre la guerra Israel-Hamas: Análisis de noticias falsas difundidas en España. *Estudios sobre el Mensaje Periodístico*, 29(2), 951-970., cuya cobertura circula a través de plataformas como Telegram, X (antes Twitter) o incluso el viejo Facebook.

50

Estos escenarios muestran que el fenómeno persiste, aunque ahora se ve amplificado por las dinámicas digitales, y de la mano de influencers, usuarios anónimos o actores de los que muchas veces desconocemos sus intereses, pero que replican y distorsionan en tiempo real las imágenes y relatos de la confrontación.

2. ¿Qué entendemos hoy por posverdad? De invención léxica a término de diccionario

El Diccionario de la Real Academia Española definió posteriormente posverdad con una formulación tan precisa como inquietante: *“Distorsión deliberada de una realidad, que manipula creencias y emociones con el fin de influir en la opinión pública y en actitudes sociales”*. De allí que resulte pertinente articular esta definición con la del Diccionario Oxford y a partir de ambas, descomponer el concepto para extraer cuatro ingredientes que en conjunto, permiten describir su verdadera naturaleza. Este ejercicio facilita comprender por qué la posverdad resulta tan difícil de enfrentar en el marco de una democracia que privilegia la libertad de expresión.

Del ejercicio anterior obtenemos entonces cuatro elementos o ingredientes claves, cada uno de distintas naturalezas, así:

1. **Acción:** *“distorsión deliberada de una realidad”*; Este es el punto de partida del fenómeno y nos indica que detrás existe un acto consciente que consiste en alterar hechos verificables para producir una versión distinta de los acontecimientos.

2. **Intención:** *“manipular creencias y emociones”*; Este elemento revela que quienes ejecutan esa acción buscan incidir en la percepción subjetiva más que en la capacidad crítica de contrastar datos, recurriendo para ello a medios ilegítimos y arteros.

3. **Propósito:** *“influir en la opinión pública y en las actitudes sociales”*; Aquí se señala la finalidad del fenómeno, que es generar un espacio en el que la manipulación se traduzca en resultados concretos. Resultados que pueden expresarse en conductas políticas, decisiones electorales o incluso en orientaciones culturales, de modo que la distorsión de la realidad no permanezca en el plano discursivo, sino que se materialice en acciones colectivas.

4. **Circunstancia:** *“las personas responden más a sentimientos y creencias que a los hechos”*; Este escenario facilita la eficacia social de la distorsión, pues otorga a la emoción y a la subjetividad un peso mayor que al dato verificable, creando un terreno fértil para la expansión de la posverdad.

51

Esta disección del significado de la posverdad permite entrever por qué, en la era de las telecomunicaciones, la desinformación circula con una libertad sin precedentes. Precisamente porque jugar con estos elementos dentro del ecosistema de la Web 2.0, donde abundan los recursos multimedia y donde el anonimato, o la “virtud” de la multiplicidad de un mismo sujeto a través de varias cuentas, se encuentran al alcance de cualquiera, no solo facilita la propagación de la mentira, sino que la amplifica y la replica de manera exponencial, y el resultado es un escenario en el que los hechos compiten en clara desventaja frente a emociones y creencias, entre “like” y “reposteos”. A continuación, revisemos cada ingrediente por separado:

2.1. **Acción:** “distorsionar deliberadamente una realidad”.

Distorsionar la realidad no es, en sí mismo, un acto ilegítimo; en el arte y la literatura constituye, de hecho, un recurso creativo consagrado. Picasso deformaba los rasgos humanos para capturar lo esencial de la experiencia, y Juan Gabriel Vásquez recuerda que la novela busca “decir solo lo que la novela puede decir”. En ese sentido, toda ficción transforma la realidad para comprenderla mejor, pues en el terreno de

lo simbólico, la distorsión es una herramienta válida, necesaria, para explorar la condición humana, comunicar emociones y construir sentidos colectivos.

El problema aparece cuando esa distorsión deja de ser estética o retórica y se convierte en un mecanismo intencional de manipulación. No hablamos aquí de imaginar mundos posibles ni de reinterpretar el pasado de forma poética, sino de alterar hechos verificables para instalar narrativas que compiten con la verdad y la desplazan.

Así, lo que en una pintura o en una novela es legítimo, en materia política o en la discusión pública puede resultar en un acto embustero o manipulador, y ejemplos abundan en las redes sociales, como los que aseveran que la Tierra es plana, que las vacunas contra el COVID-19 contienen microchips o son para cambiar nuestra genética humana, que hay gobernantes que son “enviados de Dios” para salvar a una nación, o las famosas conspiraciones sobre que los miembros de la realeza no son humanos si no extraterrestres “reptilianos”, y hasta asuntos de celebridades a las que se les considera inmortales. En todos estos casos, la distorsión se proyecta hacia la sociedad con el fin concreto de confundir, inducir o alterar percepciones.

La Real Academia Española define distorsionar como la acción de presentar o interpretar hechos e intenciones “*deformándolos de modo intencionado*”, la definición implica voluntad de engaño, y en democracia, el engaño masivo deteriora la deliberación pública impidiendo el debate de argumentos, y este es el punto de inflexión, mientras la ficción exagera, romantiza, dramatiza y juega con la realidad para reinterpretarla o dotarla de sentido, la posverdad la deforma para manipularla, porque su objetivo no es la belleza, el entretenimiento, la reflexión o la imaginación, sino provocar en el receptor una percepción errónea, construida deliberadamente para condicionar su conducta.

El desafío democrático que emerge de este primer ingrediente de la posverdad es complejo, pues una democracia liberal descansa en la libertad de expresión, y esa libertad incluye la posibilidad de exagerar, satirizar, “reimaginar” o criticar. Sin embargo, cuando la distorsión deliberada se convierte en estrategia política, el riesgo es que el contenido del debate público se torne vacío, superfluo o espurio, desplazando el valor de la verdad, de los datos y de los argumentos. Esto ocurre con frecuencia cuando se simplifican en exceso las soluciones a los grandes problemas que padece una sociedad, sustituyendo la deliberación crítica por eslóganes emocionales.

En síntesis este punto evidencia que el núcleo operativo de la posverdad consiste en actuar como un mecanismo que, bajo la apariencia de libertad de expresión, degrada la confianza ciudadana y convierte a la democracia (en donde la libre expresión es un principio esencial) en un terreno fértil para la manipulación y el engaño.

2.2. **Intención:** “manipular creencias y emociones”.

La manipulación de creencias y emociones no siempre constituye un acto ilegítimo, de hecho, forma parte de la vida cotidiana en una democracia. Muchas campañas de salud pública, por ejemplo, apelan simultáneamente a la razón y a la emoción para fomentar conductas preventivas, como lo es, mostrar la imagen de un pulmón dañado en una caja de cigarrillos, eso no es un engaño, es un recurso de persuasión legítimo, que recurre a un lenguaje emocional orientado a conectar con miedos y esperanzas reales, con el fin de generar conciencia y guiar decisiones en beneficio colectivo.

Para examinar este aspecto con mayor precisión, conviene recordar una lección que suele enseñarse en las primeras clases de filosofía en el bachillerato. En su célebre obra “La Retórica”, Aristóteles estableció que la persuasión descansa en tres pilares, que a saber son: el ethos, la credibilidad del orador, el pathos, que es la disposición emocional del auditorio, y el logos que es la razón y prueba; estos tres pilares en conjunto, funcionan como criterios para establecer la validez de un determinado discurso, y este marco clásico, nos recuerda que la apelación emocional es legítima siempre que se sustente en hechos y se oriente a fines públicos; De hecho, aún hoy sus categorías siguen siendo un referente para distinguir entre persuasión y manipulación en el espacio democrático.

En efecto, Aristóteles enseñaba que el pathos, es decir, la apelación a las emociones, es tan válido como el logos (la razón) o el ethos (la credibilidad), y en este sentido, influir en el mundo emocional de los ciudadanos constituye un recurso válido para movilizar decisiones colectivas y generar consensos.

El quiebre está cuando la influencia deja de ser transparente, cuando deja de ser persuasión y se convierte en una manipulación dolosa o intencionada, dado a que mientras la persuasión implica exponer argumentos para ser evaluados, la manipulación oculta las intenciones y busca explotar vulnerabilidades emocionales, y allí vuelve a entrar en juego la posverdad, pues la mentira es fabricada para insertarse en las creencias preexistentes de las personas, en sus sesgos, para poderse

amplificar a través de emociones intensas.

En razón de lo anterior, el desafío democrático de este segundo ingrediente radica en que la manipulación de emociones y creencias erosiona la confianza en el debate público, pero al mismo tiempo, su presencia en una democracia no puede ser proscrita sin amenazar la libertad de expresión. Influnciar emociones es parte natural de la comunicación política y ciudadana; lo problemático surge cuando esa influencia abandona la transparencia y se convierte en manipulación deliberada.

Adicionalmente, esa dificultad se agrava en un entorno de redes sociales donde el mitin político parece permanente, entre memes, chistes, críticas y quejas, y en el que periodistas, políticos, influencers y ciudadanos comunes comparten el mismo espacio digital, interactuando de tú a tú dentro del patrón de relación entre poder, medios y ciudadanía.

54

En ese escenario, como ya lo advertía Tesich en 1992, la verdad tiende a ceder ante las creencias o emociones, y de ahí surgen dos preguntas clave, ¿es realmente posible diferenciar la influencia legítima de la manipulación ilegítima en un entorno mediático donde la verdad y la mentira se difunden con distinta velocidad y alcance? ¿Y con qué legitimidad podría alguien pretender fijar esa frontera, cuando lo que está en juego son precisamente las creencias que nos definen como individuos y comunidades? Incluso a comienzos del siglo XX, Mark Twain dejó constancia de esta tensión en su Autobiografía, donde escribió *“How easy it is to make people believe a lie, and how hard it is to undo that work again!”* (Twain, 2013), es decir *“Qué fácil es hacer que la gente crea una mentira, y qué difícil es deshacer ese trabajo una vez hecho.”*

2.3. Propósito: influir en la opinión pública y en actitudes sociales

Para el examen de este tercer ingrediente de la posverdad, el de tipo propósito, *“influir en la opinión pública y en las actitudes sociales”*, resulta pertinente apoyarse en tres referentes clásicos del pensamiento democrático, Jürgen Habermas, Norberto Bobbio y Robert Dahl. Cada uno, desde su marco conceptual, ha mostrado que el debate público, la competencia entre propuestas y la participación ciudadana no son adornos del sistema, sino pilares que hacen posible la democracia misma; Influir en la opinión pública o sus actitudes sociales no solo es válido en una democracia, de hecho, es su razón de ser, es lo que permite participación y constituye en gran medida uno de sus pilares

fundamentales, ¿qué sería de nuestra democracia si no se permitiera el debate público entre candidatos?, en efecto sería simplemente una dictadura (Habermas, 1994).

Tanto el debate político como las campañas electorales, las protestas ciudadanas o las campañas institucionales buscan orientar la conducta colectiva hacia determinadas decisiones; el llamado a marchar por una causa, la convocatoria a una manifestación en contra o a favor de una medida, o la promoción de una reforma, son distintas formas de persuasión legítima que procuran incidir en una ciudadanía activa para alcanzar mayorías (Bobbio, 1984). Ciertamente, la vitalidad de una democracia se sostiene precisamente en la circulación de mensajes que compiten por atraer voluntades, generar consensos entre iguales o moldear conductas, y en ese sentido, votar, marchar, apoyar una reforma o rechazar una medida son actos sociales que nacen de esa influencia legítima ejercida en el espacio público (Dahl, 1992).

El problema surge cuando esa influencia abandona la transparencia y recurre a la falsedad como herramienta de incidencia, pues no es lo mismo convencer a una comunidad de que debe ahorrar agua con datos verificables sobre el cambio climático, que movilizarla mediante mensajes manipulados que tergiversan la evidencia científica. En el ámbito político, por ejemplo, el quiebre es aún más delicado, basta recordar cómo, en noviembre de 2019, cadenas de audios y mensajes falsos que circularon por WhatsApp fueron suficientes para sembrar pánico en Bogotá y llevar a la imposición de un toque de queda, un episodio en el que el miedo y la confusión ciudadana derivaron en caos colectivo (Caballero, 2020), caso que demuestra que el miedo incidió torciendo la realidad para dirigir conductas sociales.

El trasfondo académico confirma la gravedad del asunto. Benkler, Faris y Roberts, en *Network Propaganda* (2018), demostraron cómo las redes digitales de información no solo amplifican narrativas partidistas, sino que también reordenan la agenda pública hacia problemas cargados de emociones, relegando los hechos; Así, el propósito de influir se transforma en que ya no se trata de persuadir con argumentos, sino de imponer realidades alternativas que condicionan el comportamiento colectivo.

2.4. Circunstancia: las personas responden más a sentimientos y creencias que a hechos

Ninguno de los ingredientes examinados anteriormente sería posible sin este cuarto y último elemento de la posverdad, señalado

en la definición del Diccionario Oxford, un rasgo propio de la naturaleza humana según el cual “*las personas responden más a sentimientos y creencias que a hechos*” (Oxford University Press, 2016). Esto no es en sí mismo un defecto, es parte de nuestra condición humana y un hecho reconocido desde la antigüedad.

En democracia, esta inclinación puede incluso ser positiva, porque la solidaridad frente a una tragedia, la indignación ante la injusticia o la esperanza de un cambio político son motores que mueven a las sociedades a actuar cuando los hechos, por sí solos, resultarían insuficientes (Nussbaum, 2014). El quiebre ocurre cuando esta predisposición natural se convierte en un terreno explotado sistemáticamente para manipularnos, y al parecer la comunicación política digital lo ha entendido con claridad, no basta con presentar datos, hay que construir narrativas que despierten miedo, rabia o entusiasmo (Amorós García, 2018).

56

Para comprender de manera sencilla lo que significa afirmar que las personas responden más a sentimientos y creencias que a hechos, pensemos en un ejemplo cotidiano y neutral. Imagine, por un instante, un alimento que usted jamás probó porque estaba convencido de que no le agradaría, quizás por su olor, su textura o la creencia de que le haría daño a la digestión, en consecuencia, cada publicación negativa que aparecía en redes sociales sobre ese alimento usted la compartía, convencido de que nadie más debía consumirlo ni repetir aquella “mala experiencia”. Se unió a grupos en los que esa percepción era reafirmada una y otra vez, mediante videos cortos, memes, stickers y comentarios que reforzaban la convicción sin necesidad de ponerla a prueba.

Durante años lo que lo llevó a rechazarlo no fue el hecho en sí, sino la fuerza de una creencia sostenida emocionalmente, hasta que, por azar o curiosidad, un día lo probó (quizás porque no sabía que ese ingrediente estaba en su plato o porque alguien lo incluyó en una receta) y descubrió que, lejos de ser desagradable, le fascinaba al punto de incorporarlo a su dieta.

Lo interesante de este ejemplo es que lo que había determinado su conducta no era la realidad, sino una emoción que nunca había contrastado con la experiencia, así mismo ocurre también en la esfera pública, nuestras predisposiciones emocionales y cognitivas funcionan como filtros que nos protegen, pero también pueden cerrarnos a escuchar o considerar alternativas (Pariser, 2011). Este mecanismo, que en lo individual es natural, se convierte en terreno fértil para la explotación política cuando se refuerza con memes, ironías, chistes o imágenes virales que, amplificadas en redes sociales, radicalizan posiciones

y reducen la posibilidad de diálogo.

3. La posverdad en la era digital (redes sociales, narrativas virales y filtro burbuja)

En la era de la posverdad, y a estas alturas del año 2025, el riesgo ya no está solo en la circulación de noticias falsas, sino en la propia arquitectura comunicativa contemporánea. Hoy las redes sociales, consolidadas con millones de usuarios, han sido diseñadas para premiar lo que genera reacciones rápidas y emocionales; sus algoritmos construyen burbujas de información que refuerzan creencias preexistentes y alimentan sesgos cognitivos. A este panorama se suma lo que Daniel Kahneman, en su obra *“Thinking, Fast and Slow”* (Pensar rápido, pensar despacio), explicó sobre el modo en que los seres humanos procesamos la información a partir de dos sistemas de pensamiento. El Sistema 1 es rápido, intuitivo y emocional, procesa información de manera automática, basada en asociaciones y atajos mentales; es el que nos permite reaccionar con inmediatez, pero también el más propenso a sesgos y errores. El Sistema 2, en contraste, es lento, racional y analítico, exige esfuerzo cognitivo, revisión crítica y contraste de datos.

57

En la práctica, sin embargo, gran parte de nuestras decisiones cotidianas, y también políticas, se toman bajo el dominio del *Sistema 1*, mientras el *Sistema 2* rara vez interviene para corregir esos atajos, de ahí que nuestras emociones actúen como filtros inmediatos de la realidad, mientras la razón crítica llega tarde o no alcanza a contrarrestar narrativas falsas que ya han sido instaladas en la conversación pública.

Es así como las redes sociales moldean nuestra percepción de la realidad, porque muchos ya no nos damos la oportunidad de ver el mundo con nuestros propios ojos, sino que lo hacemos a través de la perspectiva que otros construyen para nosotros y que se ve amplificada por las plataformas digitales; y en medio de esa inmediatez y de millones de contenidos producidos por millones de usuarios, esperar que cada uno de nosotros actúe como un “verificador” constante de contenido es por poco ingenuo, y de paso condena a la democracia a una debilidad permanente.

3.1. Las Redes sociales (Web 2.0)

La Web 2.0 le otorgó a los usuarios de la red un rol activo en la producción, distribución y validación de contenidos, pasamos de una web plana y estática a una en la que todos, de algún modo, co-construimos, de hecho, Dan Gillmor sintetizó esta idea con la frase *“We, the media”*.

En este contexto, redes sociales como YouTube, Facebook, WhatsApp, Instagram, TikTok y X/Twitter se han izado como referentes globales que ya no funcionan solo como espacios de ocio o de intercambio cotidiano, sino como verdaderos escenarios de debate y movilización política permanente, donde la interacción, la publicidad, la participación política y el consumo informativo se entrelazan de manera constante.

Los datos confirman la magnitud del fenómeno de las redes sociales, por ejemplo, en América Latina, según el informe *Digital 2025 de We Are Social & Meltwater*, el 86 % de la población es usuaria activa de estas redes, lo que equivale a más de 460 millones de personas conectadas que comparten y consumen contenidos. En España, la cifra asciende al 91 % de la población, con 43 millones de usuarios activos (We Are Social & Hootsuite, 2025); y en Estados Unidos supera los 246 millones de personas, lo que representa cerca del 73 % de su población, de acuerdo con el *Pew Research Center* (2024).

58

En cuanto al tiempo de uso, el promedio mundial es de 2 horas y 23 minutos diarios, sin embargo, en América Latina la cifra es sensiblemente mayor y supera ampliamente las 3 horas en promedio de acuerdo con Hootsuite. Estas estadísticas no solo dimensionan el alcance del fenómeno, sino que muestran hasta qué punto la vida social y política contemporánea se desarrolla en entornos digitales que ya no son marginales, sino constitutivos de la esfera pública. Bajo ese panorama, el tipo de contenido más compartido se orienta hacia lo visual, imágenes y videos breves de allí que las más usadas sean TikTok, Reels, YouTube Shorts e Instagram, seguidos por memes y mensajes de texto cortos, todos diseñados para generar una reacción emocional inmediata (Statista, 2024).

De este modo, las redes sociales han pasado a compartir y reconfigurar un rol que antes parecía exclusivo de los medios tradicionales, es decir, el de construir realidades sociales, orientar la opinión pública y abrir la puerta a fenómenos propios de la posverdad (Castells, 2009) y si a esto le sumamos, como muestran los datos ya enunciados, que estos espacios digitales se han convertido en parte central de la vida cotidiana, donde se refleja la conversación pública y, al mismo tiempo, se la moldea de manera constante a través de clics, likes, compartidos e interacciones, vemos que se configuran entornos que contribuyen a orientar creencias y sentimientos que, en la práctica, pueden llegar a pesar más que los hechos en la construcción de legitimidad política (Gillmor, 2004).

3.2. El Filtro burbuja

Apoyándonos en la conferencia de Eli Pariser “*Fake news, los filtros burbuja y la batalla por la verdad*” (Fundación Telefónica, 2017) y disponible en YouTube, el filtro burbuja puede entenderse como un efecto colateral de la personalización algorítmica, es decir, un mecanismo que prioriza lo que cada usuario ve según sus interacciones previas, gustos y redes, excluyendo o invisibilizando aquello que contradice sus creencias, de allí que lo que en un principio fue creado para optimizar la experiencia de navegación en plataformas digitales, termina convirtiéndose en un recurso que refuerza convicciones y reduce la exposición a perspectivas divergentes sobre cualquier tema.

La dificultad radica en que estos algoritmos no fueron creados para valorar la veracidad de los contenidos, ni para distinguir si los datos, videos o audios compartidos son auténticos o responden a intenciones manipuladoras, su único objetivo es maximizar la permanencia del usuario dentro de la red social, y es que, aunque estas plataformas aparentan ser gratuitas, su lógica de negocio se sustenta en la publicidad y en la explotación de datos; de ahí la necesidad de retener a los usuarios ofreciéndoles aquello que más coincide con sus preferencias. El resultado es una constante exposición a mensajes de todo tipo (comerciales, culturales y políticos) que buscan modelar tendencias, percepciones y emociones, generando un ecosistema informativo altamente personalizado pero cada vez más cerrado.

Es por eso que un contenido falso con alta interacción puede superar en visibilidad y viralidad a uno verdadero, hasta convertir la falsedad en una narrativa permanente, diseñada a la medida de los gustos y creencias de quien la consume (Amorós García, 2018). El resultado es un escenario en el que casi no existe posibilidad de contrarrestar, debatir, matizar o construir conciliaciones públicas.

En este sentido los algoritmos son el dictador de nuestro tiempo, en la práctica establecen qué podemos ver, escuchar, leer y debatir, así que mientras el comportamiento de estos algoritmos no sean matizados o modificados, esta personalización excesiva no garantizará un consumo de información libre, sino que al mejor estilo de una dictadura digital, seguirá controlando una determinada línea editorial a la que cada usuario esta hoy día sometido, censurando y depurando contenidos que en su entender como herramienta no nos interesan.

Así las cosas, en la actualidad cada ciudadano global que habita una red social específica está sesgado por un ecosistema informativo

propio, diseñado a partir de sus gustos y preferencias, que lo mantiene aislado de perspectivas divergentes e impide el contraste con otras miradas, contexto que va en contra vía de lo que advertía Norberto Bobbio, que el contenido mínimo de un Estado democrático exige garantías de libertad, pluralidad de partidos, elecciones periódicas y, sobre todo, un debate libre entre las partes, situación que no favorecen las redes sociales que tienden a radicalizarnos.

3.3. Las narrativas virales de carácter emocional.

Las narrativas virales suelen simplificarse a decir que son en sí mismas la posverdad, pero como hemos mostrado a lo largo de este artículo, en realidad estas son apenas un recurso dentro de un fenómeno más amplio, de hecho como ya lo hemos dejado claro, la posverdad constituye un contexto, un clima cultural, una época en el que la mentira circula con eficacia, mientras que los mensajes emocionales y las noticias falsas actúan como mecanismos de manipulación y distorsión.

60

Estas narrativas virales, hoy tienen la capacidad de operar bajo la forma de memes con carga ideológica, audios manipulados o videos sintéticos generados con inteligencia artificial, y su eficacia descansa en mecanismos conocidos como el sesgo de confirmación, la activación emocional inmediata y la inserción en narrativas preexistentes (Albo Coss, 2024). Su peligrosidad aumenta cuando se visten con el ropaje de los medios de comunicación tradicionales, es decir, cuando se presentan como noticias serias con apariencia de investigación periodística, pero en realidad todo en ellas es falso, y lo hacen porque precisamente el consumidor espera de sus medios información confiable y veraz (Hurtado, 2024).

Este riesgo se intensifica en la actual era de la inteligencia artificial generativa, ya que las herramientas digitales que antes exigían altos conocimientos técnicos y una gran dedicación para producir fragmentos convincentes hoy están al alcance de cualquiera. Como resultado, los montajes rudimentarios que eran realizados antes por seguidores entusiastas, usuarios con pocos recursos o aficionados sin formación especializada, hoy han sido reemplazados por piezas audiovisuales pulidas, realistas y altamente persuasivas. Ya no hablamos solo de noticias falsas, hoy día los periodistas y entrevistados pueden ser suplantados con deepfakes, y escenarios completos —lugares y contextos— pueden recrearse con verosimilitud absoluta, de hecho, en el año 2024 la UN-ESCO (2024) advirtió sobre la proliferación de deepfakes, la microsegmentación y las campañas de desinformación automatizada, elementos que intensifican los riesgos para la integridad democrática, sobre todo

en los ciclos electorales.

Para finalizar, imaginemos por un instante cómo habrían sido episodios como Watergate, Irán-Contra o la Guerra del Golfo (los mismos que Tesich utilizó en 1992 para formular su silogismo de la posverdad) si hubieran ocurrido en un ecosistema influido o determinado por redes sociales, filtros burbuja y narrativas virales.

Es probable que la prensa no hubiese contado con el mismo margen para marcar la agenda ni para revelar la totalidad de los hechos (como pasa hoy día), la verdad de esos casos podría haberse visto sepultada o diluida entre rumores, memes, audios manipulados y videos falsos; Richard Nixon habría tenido a su disposición una maquinaria de bots y seguidores digitales capaces de desviar la atención pública, como ocurre hoy cuando un gobierno utiliza su ejército de comunicadores en redes sociales para desorientar la opinión, inventar crisis externas o desplazar la mirada colectiva frente a problemas internos; El debate público, en consecuencia, se habría trasladado al terreno digital, donde las emociones pesan más que los datos y los usuarios terminamos sesgados por sus propias creencias.

61

La diferencia es sustantiva, lo que Tesich describió entonces como una anomalía de la democracia estadounidense se ha convertido hoy en una condición estructural de la esfera pública global, la posverdad ya no depende únicamente de gobiernos que encubren o de ciudadanos que rehúyen la verdad, sino de un ecosistema digital que multiplica narrativas emocionales hasta transformarlas en realidades paralelas, y en este entorno hiperconectado, la manipulación no opera por censura, sino por adhesión voluntaria a lo verosímil.

Más de treinta años después, la advertencia de Tesich se confirma, vivimos instalados en un tiempo posterior a la verdad, y el desafío contemporáneo no es solo tecnológico sino cultural, debemos recuperar la conversación directa, el debate informado y la valoración de los hechos verificables. Ello exige desconectarnos de la superficialidad de las redes sociales, restituir el valor de la experiencia propia y privilegiar las fuentes que se sostienen en reputación, ética y años de credibilidad, evitando la tentación de informarnos a través de influenciadores, la defensa de la democracia pasa por volver a confiar en aquellos espacios de comunicación que tienen mucho que perder si traicionan la verdad.

Ojalá no llegue el día en que como el personaje Neo de la película *Matrix* de 1999, tengamos que escoger entre la píldora roja o azul, para escapar de la *Matrix* que las redes sociales están construyendo a

nuestro alrededor; o que necesitemos de un Morfeo que nos confronte con la realidad de los hechos y los datos, que hoy cada vez más se ven sacrificados por el lente distorsionador de los entornos digitales.

Referencias

- Albo Coss, Ú. (2024). Meme. En G. Sosa Hernández & K. Gilas (Coords.), *¿Cómo estudiar la comunicación política? Nuevas miradas a los conceptos centrales* (pp. 259-264). Universidad Nacional Autónoma de México.
- Amorós García (2018). *Fake News La verdad de las Noticias Falsas*. Editorial Plataforma.
- Bernstein, C., & Woodward, B. (1974). *All the president's men*. Simon & Schuster. <https://www.thedailystar.net/sites/default/files/upload-2014/gallery/pdf/atpm.pdf>
- Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press. <https://library.oapen.org/viewer/web/viewer.html?file=/bitstream/handle/20.500.12657/28351/9780190923624.pdf?sequence=1&isAllowed=y>
- Bobbio, N. (1984). *El futuro de la democracia*. Fondo de Cultura Económica. Título original: Il futuro della democrazia (6/ 1984, Giulio Einaudi Editore, s.p.a, Turín .(Página 25 en adelante) disponible en <https://fundamentoscpuba.files.wordpress.com/2020/04/bobbio-norberto-2012-1984-el-futuro-de-la-democracia.pdf>
- Caballero, C. (2020, agosto 24). Persiguiendo fantasmas: el toque de queda de noviembre. *Razón Pública*. <https://razonpublica.com/persiguiendo-fantasmas-toque-queda-noviembre/>
- Castells, M. (2009). *Comunicación y poder* (M. Hernández, trad.; Edición electrónica, 2018). Madrid: Alianza Editorial. <https://storage.googleapis.com/warehouse-europe-west3-prod-assets/b4140915-9fc5-4952-818d-ec3335c9ed56/9788491813811/f97e46afa21ff5fe6b4e4315803d5e10.pdf>
- Costa, J. (2021, 15 de marzo). La columna de Joan Costa: ¿Qué es la verdad? *Experimenta*. <https://www.experimenta.es/a-costa-mia/la-columna-de-joan-costa-que-es-la-verdad/>
- Dahl, R. A. (1992). *La democracia y sus críticos* (L. Wolfson, Trad.). Ediciones Paidós Ibérica. (Obra original publicada en 1989). <https://iuscol.com/wp-content/uploads/2022/08/Dahl-Democracia-y-sus-criticos-6a-part.pdf>
- Digital 2025 - We Are Social Spain: <https://wearesocial.com/es/blog/2025/02/digital-2025/>
- Fundación Telefónica Madrid. (2017, 19 de junio). *Eli Pariser: Fake news, los filtros burbuja y la batalla por la verdad* (Español) | #TechSoci-

- ety [Video]. YouTube. <https://www.youtube.com/watch?v=D6QM-d9UbW2Q>
- Gillmor, D. (2004). *We the media: Grassroots journalism by the people, for the people*. Sebastopol, CA: O'Reilly Media. <https://archive.org/details/tails/wethemediagillmor>
- Habermas, J. (1994). *Historia y crítica de la opinión pública: La transformación estructural de la vida pública* (A. Doménech, Trad.; con la colab. de R. Grasa; prefacio de la reed. alemana de 1990 por F.J. Gil Martín; rev. bibliográfica por J. Romaguera). Editorial Gustavo Gili. (Obra original publicada en 1962). <https://archive.org/details/habermas-j.-historia-y-critica-de-la-opinion-publica-1994/page/n3/mode/2up>
- Hurtado, L. (2024). Fake news. En G. Sosa Hernández & K. Gilas (Coords.), *¿Cómo estudiar la comunicación política? Nuevas miradas a los conceptos centrales* (pp. 201-206). Universidad Nacional Autónoma de México.
- Kellner, D. (1992). *The Persian Gulf TV war*. Boulder, CO: Westview Press. https://archive.org/details/persianguftvwar0000kell_l7c9/page/n7/mode/2up
- Kornbluh, P., & Byrne, M. (1993). *The Iran-Contra Scandal. The Declassified History*.
- Oxford University Press (2016). *Word of the Year: Post-truth*. Oxford Dictionaries. <https://languages.oup.com/word-of-the-year/2016/>
- Kreab (2022). La influencia de los medios de comunicación y las redes sociales en el conflicto Rusia-Ucrania. *Kreab Explains*. <https://kreabexplains.es/financiero/la-influencia-de-los-medios-de-comunicacion-y-las-redes-sociales-en-el-conflicto-rusia-ucrania/>
- Massoni, S. (2019). Teoría de la Comunicación Estratégica Enactiva e Investigación Enactiva en Comunicación: aportes desde Latinoamérica a la democratización de la vida cotidiana. *Chasqui. Revista Latinoamericana de Comunicación*, 141, 237–256. <https://revista-chasqui.org/index.php/chasqui/article/view/4074/3227>
- Nussbaum, M. C. (2014). *Emociones políticas: ¿Por qué el amor es importante para la justicia?*. https://proassetspdldcom.cdnstatics2.com/usuarios/libros_contenido/arxius/29/28428_Las_emociones_politicas.pdf
- Oxford University Press. (2016). *Word of the Year: Post-truth*. Oxford Dictionaries. <https://languages.oup.com/word-of-the-year/2016/>
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. https://youtu.be/qHXBCoRo1aE?si=2HcuVQTn_f3Ef_sq
- Pérez, R. A. (2014). La estrategia como campo de estudio: ¿Tenemos ya un nuevo paradigma? *Revista Mediterránea de Comunicación*, 5(2), 9–21. <https://www.mediterranea-comunicacion.org/article/view/2014-v5-n2-la-estrategia-como-campo-de-estudio-tene->

[mos-ya-un-nue](#)

- Robinson, M. J. (1974). Public opinion during the Watergate crisis. *Communication Research*, 1(3), 363–381. <https://doi.org/10.1177/009365027400100404>
- Robinson, P. (2002). *The CNN effect: The myth of news, foreign policy and intervention*. Routledge. <https://www.guillaumenicaise.com/wp-content/uploads/2013/10/the-cnn-effect.pdf>
- Ross, AS, y Rivers, DJ (2017). *Culturas digitales de participación política: Memes de internet y la deslegitimación discursiva de los candidatos presidenciales estadounidenses de 2016*. Discurso, contexto y medios, 16, 1-11. <https://doi.org/10.1016/j.dcm.2017.01.001>
- Ruiz-Baena, M. (2023). Desinformación sobre la guerra Israel-Hamas: Análisis de noticias falsas difundidas en España. *Estudios sobre el Mensaje Periodístico*, <https://revistas.ucm.es/index.php/ESMP/article/view/95088/4564456569631>
- Statista. <https://es.statista.com/sectores/1176/internet/>
- Taylor, P. M. (1992). *War and the media: Propaganda and persuasion in the Gulf War*. Manchester University Press. https://books.google.co.bw/books?id=V9tRAQAAIAAJ&printsec=frontcover&source=gbs_atb#v=onepage&q&f=false
- Tesich, S. (1992, January 6). A government of lies. *The Nation*, 254(1), 12–14. <https://archive.org/details/steve-tesich-government-of-lies-article>
- The New Press. <https://archive.org/details/irancontrascanda00korn/page/n5/mode/2up>
- Thomas, C. (1978). Presidential accountability since Watergate. *Political Science Quarterly*, 93(4), 605–627. <https://doi.org/10.2307/2149804>
- Twain, M. (2013). *Autobiography of Mark Twain: Volume 2 (The complete and authoritative edition)*. University of California Press. (Original work published 1906). https://nibmehub.com/opac-service/pdf/read/Autobiography%20of%20Mark%20Twain-%20Volume%202%20_%20the%20complete%20and%20authoritative%20edition.pdf
- UNESCO (2024). *Libertad de expresión, inteligencia artificial y elecciones. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura*. https://unesdoc.unesco.org/ark:/48223/pf0000393473_spa

Presentación de la segunda jornada

Nos encontramos, sin lugar a dudas, en el umbral de una crisis que trasciende lo meramente político y lo económico: se trata de una crisis epistémica. Y es importante dejarlo claro: no es una crisis de información en sí misma, ¡al contrario!, sino de verosimilitud, de la confianza en el método, o en las intenciones subyacentes a cualquier “verdad” que se nos presente. En la era digital, las fuentes son intrínsecamente diversas, sí; sin embargo, la credibilidad se ha puesto en juego, gravemente. Ya no existe una única institución legítima que avale —que investigue y respalde— la información como solía ocurrir en el pasado cercano.

Este libro, que es el resultado directo de nuestro congreso anual FUTURIBLE 2025, nace de la urgente necesidad de profundizar, nombrar y dismantelar el fenómeno que está redefiniendo los cimientos de nuestra sociedad: la desinformación. No hablo, en absoluto, de un simple desacuerdo ideológico; ni siquiera de la mentira fácil que siempre ha acompañado a campos complejos como la política. Por supuesto, es sabido que muchos regímenes, a lo largo de la historia, abusaron del control sobre la información en diversas naciones. Lo que hoy enfrentamos es distinto: hablo de un asalto estratégico y sistemático, con una sincronía global, dirigido directamente a nuestra capacidad colectiva de discernir lo real.

65

Durante décadas, la promesa de la tecnología fue la de un mundo más interconectado y, por ende, automáticamente más informado. Pero esa utopía, la de la polis digital donde el conocimiento fluiría libremente, se ha disuelto, tristemente, en un auténtico diluvio de ruido y falsedad. La información hoy no solo se democratizó en su acceso, sino también en su producción, y con esa democratización, es inevitable reconocerlo, la capacidad de engañar se ha industrializado.

El término “fake news”, tan manido en el debate público, resulta, sencillamente, insuficiente. Lo que enfrentamos es una sofisticada taxonomía de la falsedad, donde el contenido puede ser manipulado, sacado de contexto o directamente fabricado, apelando no a la razón fría, sino a la reacción visceral. Las plataformas que se crearon para conectarnos han, paradójicamente, encerrado en cámaras de eco, ecosistemas algorítmicos donde la verdad se mide no por su evidencia, sino por su resonancia emocional y, ¿por qué no decirlo?, por su capacidad cuantitativa y vertiginosa de replicarse (o quizás deberíamos decir, de reproducirse). Y los datos son contundentes: en el caso de las redes sociales, las fake news superan hasta seis veces la velocidad de propagación de

las noticias reales, según el estudio seminal del MIT publicado en The spread of true and false news online.

Esta corrosión tiene consecuencias que van mucho más allá del timeline de nuestras redes. El primer efecto, y el más tangible de todos, es la destrucción del capital de confianza —ese activo invisible que, al fin y al cabo, es lo que sostiene la convivencia civilizada—. La sistemática erosión de la reputación de expertos, científicos e instituciones nos deja a la deriva, incapaces de ponernos de acuerdo siquiera en los hechos fundamentales que requerimos para tomar decisiones colectivas. Y esta desconfianza se traduce, inevitablemente, en violencia. Rumores digitales se han convertido en la mecha que enciende actos de terrorismo estocástico y linchamientos en el mundo físico, lo cual demuestra que la línea entre el odio digital y el daño tangible es cada vez más difusa. Pero no podemos, bajo ninguna circunstancia, caer en la desesperanza. Este panorama no es un destino ineludible; es más bien una encrucijada, un desafío mayúsculo. La Inteligencia Artificial, que hoy ostenta el poder de ser la herramienta más potente para la creación de deepfakes y la propagación de narrativas masivas, tiene la obligación moral de ser también el arma más efectiva para la defensa de la verdad. La carrera armamentista tecnológica entre los actores de la desinformación y los verificadores constituye, sin duda, el campo de batalla crucial de este siglo.

El propósito central de este libro, entonces, es ofrecer una hoja de ruta clara para la gobernanza digital que esta crisis nos exige. Es una solución que, lo sabemos, no vendrá de un único actor. Requerirá que los Estados regulen inteligentemente; que las grandes plataformas asuman una responsabilidad fiduciaria genuina sobre el ecosistema de la información; y, fundamentalmente, que cada ciudadano se arme con la herramienta más poderosa: la alfabetización mediática y digital y el pensamiento crítico.

Este libro, al final, es una invitación directa a la acción, a la reflexión crítica profunda sobre nuestro consumo de información y a la defensa activa de la razón y la evidencia. Es un llamado a luchar no solo por la verdad, sino por el futuro de nuestra polis digital, un futuro donde el debate se base en hechos inamovibles y no en mentiras cuidadosamente diseñadas. La batalla por la veracidad es, una batalla por la supervivencia misma de nuestra democracia.

Dona Arrieta

Directora de CINTE

MESA

Gestión de la desinformación y la credibilidad en la inteligencia artificial

3

67

Moderadora

Maritza Mosquera de Sumich

Presidenta de la Confederación Interamericana de Relaciones Públicas, CONFIARP.
Presidenta de FISEC en Panamá. msumich@gmail.com

La era digital ha transformado el periodismo y el consumo de información, enfrentando a la sociedad a una creciente manipulación mediática facilitada por la inteligencia artificial (ia). Encuestas recientes reflejan que más de la mitad de las personas sienten inseguridad para diferenciar noticias verdaderas de falsas, y un 88% manifiesta preocupación por las fake news. Entre los problemas más relevantes se destacan la baja confianza en los medios, la fatiga informativa que provoca evasión de noticias, la proliferación de discursos polarizantes y de odio, y la consolidación de la “realidad sintética”, que incluye textos, imágenes, audios y videos generados por ia difíciles de distinguir de lo real.

La tecnología, aunque puede ser usada para manipular, también ofrece herramientas para la verificación y detección de desinformación. Plataformas como gptzero, hive moderation, syntheticeye, inviid, y buscadores inversos de imágenes permiten identificar contenidos generados artificialmente, verificar ubicaciones o analizar metadatos de videos. Sin embargo, persisten riesgos éticos, como sesgos en los modelos de IA, falta de transparencia, automatización de controles que podrían derivar en vigilancia o censura, y la utilización de deepfakes con fines políticos o de acoso. por ello, se subraya la necesidad de fortalecer el periodismo basado en la búsqueda de la verdad y la verificación rigurosa de la información.

La IA también ha acelerado la producción y difusión de contenidos falsos, aumentando la dificultad para distinguir lo auténtico de lo manipulado. Deepfakes, clonación de audios y fake news se propagan rápidamente, erosionando la confianza en evidencias visuales, auditivas y textuales. este fenómeno ha generado dinámicas de violencia en línea, amplificadas por algoritmos de redes sociales, que incluyen acoso masivo, campañas de desprestigio, estigmatización de minorías y polarización política. Los efectos se sienten tanto a nivel individual como colectivo, afectando la capacidad de las comunidades para confiar en la información y entre sí.

El impacto psicológico sobre las personas que son objetivo directo de la desinformación incluye ansiedad, disminución de la autoestima, aislamiento social y revictimización continua. A nivel social, esto se traduce en desconfianza generalizada, normalización de la violencia simbólica y debilitamiento de las relaciones grupales. en conjunto, los contenidos sintéticos y la desinformación promovidos por IA no solo desafían el periodismo y la ética mediática, sino que también constituyen catalizadores de conflictos y violencia en línea, exigiendo respuestas tecnológicas, educativas y éticas para proteger la veracidad de la información y la cohesión social.

Tenemos el gusto de presentar a dos destacados estudiosos e investigadores que abordarán el tema desde aristas complementarias. el primero: desde la cuna de los mariachis, radicado en España. tema la IA y desinformación como catalizadores de violencia en línea. Después, Doctoranda Soledad Aréquez: Estrategias para abordar y enfrentar los desafíos de la desinformación y credibilidad en la IA desde la tierra del tango. con el tema: La realidad bajo amenaza: contenidos sintéticos para desinformar.

Muchísimas gracias

Inteligencia artificial y desinformación como catalizadores de la violencia digital: Un análisis estratégico-psicosocial

70

Jorge Rendón Riba

Doctor en comunicación y pensamiento estratégico por el CADEC, cuenta con un máster en comunicación social por la Universidad Complutense de Madrid y es licenciado en comunicación por la universidad iberoamericana de la ciudad de México. Ha ampliado su formación con estudios en branding y marketing en instituciones de prestigio como el instituto europeo de estudios empresariales en España, la Birmingham City University, la universidad de Oxford en Reino Unido, y en marketing digital en We are Digital, también en el Reino Unido.

Cuenta con una trayectoria profesional en comunicación, relaciones públicas, branding y estrategias digitales tanto en México como en España. Ha trabajado como consultor y asesor en el sector público y privado, liderando agencias como grupo TRZ Comunicación, Mátrika comunicación avanzada y grupo entropía. Su experiencia incluye el desarrollo de campañas para marcas como Sony, Telcel, Warner Bros, Gillette, Dior y GSK, así como asesoría a artistas, festivales, instituciones educativas y campañas políticas. Ha sido productor asociado de cine, productor de espectáculos y creador de contenidos en radio y televisión digital.

Ha impartido clases por más de 20 años en universidades de prestigio en México y España, en temas de redes sociales, marketing digital, inteligencia artificial y comunicación estratégica y actualmente en el director del área de comunicación y del área digital en la asociación española para la prevención del acoso escolar, donde desarrolla actividades relacionadas con tecnología, exposición a pantallas, ciberacoso escolar e inteligencia artificial. También es fundador de la agencia Perfect Branding y conductor del podcast “Generando confianza”, especializado en acoso y ciberacoso escolar. rendonriba@gmail.com

Introducción

La inteligencia artificial (IA) ha cambiado las reglas del juego de la desinformación y se ha convertido, en ocasiones, en un vehículo de violencia digital con efectos devastadores a niveles políticos y psicosociales. Este texto examina cómo la IA, al producir contenidos falsos pero percibidos como reales, se convierte en un catalizador de nuevas formas de violencia en el ámbito digital que van desde el ciberacoso hasta la erosión del diálogo democrático como consecuencia de la polarización política. Asimismo, analiza cómo, a través de un proceso de interacción entre algoritmos, se logra influir en las emociones de grupo y las dinámicas de poder en un contexto que integra enfoques sociológicos, psicológicos y de teoría estratégica contemporánea de la administración. Concluye con recomendaciones para la alfabetización digital, la formación de ciudadanos críticos capaces de desenvolverse en la sociedad de la información actual, la regulación ética y la resiliencia digital.

La desinformación —el proceso a través del cual se comparte información falsa intencionalmente para causar daño (Wardle & Derakhshan, 2017) o contenido manipulado deliberadamente, difundido con el objetivo de engañar al público y alcanzar fines estratégicos, políticos o económicos, según la definición del Parlamento Europeo— no es algo nuevo. Sin embargo, la aparición de la IA ha cambiado las reglas del juego: ahora no se trata solo de la circulación de rumores o noticias inventadas en ámbitos relativamente acotados, sino de la construcción de realidades artificiales con videos o audios hiperrealistas, contenidos falsos producidos por algoritmos para ser utilizados en campañas a gran escala.

Este uso de una nueva tecnología no solo distorsiona la realidad de los hechos, sino que también puede emplearse como un catalizador de la violencia digital, definida como cualquier acción deliberada que, por medio de canales digitales, genera daños a nivel psicológico, social o político a individuos o comunidades. La IA y la desinformación, desde un punto de organización social, funcionan como armas de “desestabilización política suave”, es decir, la erosión indirecta y gradual de la autoridad de un Estado en un sistema democrático. Su poder se basa en la capacidad de desgastar la confianza y dividir el tejido social sin necesidad de recurrir a la fuerza física.

En primer lugar, vamos a explicar cómo la IA potencia los efectos de la desinformación para después mostrar sus consecuencias en los ámbitos psicosocial y político; para finalizar, hablaremos de la pre-

vención de estos peligros en la era digital.

Hoy en día, la desinformación ha evolucionado de una distorsión de los acontecimientos a una fabricación de realidades paralelas que se consideran verdaderas. Uno de los factores que han contribuido a este cambio es la inteligencia artificial generativa, que produce contenidos nuevos como textos, imágenes, videos o música utilizando los patrones que ha adquirido a partir de grandes volúmenes de datos. Las herramientas actuales, o modelos de lenguaje generativo como DALL-E, GPT-4, Stable Diffusion o Midjourney, son mucho más sofisticadas que cualquier otra herramienta existente hasta ahora y posibilitan la creación de contenidos falsos con una capacidad persuasiva sin igual en la historia de la comunicación humana (Törnberg, 2023). Este cambio no significa una mera ampliación de las mentiras tradicionales; conlleva un auténtico cambio de paradigma. Este nuevo modelo tiene tres formas fundamentales:

72 - Fake news automatizadas que son modelos de lenguaje generativo capaces de crear artículos, noticias y análisis que reproducen con exactitud la forma y el estilo de los medios informativos verídicos. Estos contenidos tienen la capacidad de adaptarse a grandes escalas para diversos públicos y contextos culturales (Törnberg, 2023). La velocidad de producción es tan alta que el control humano y la verificación de hechos no pueden seguirle el paso ni detener su difusión.

- Deepfakes audiovisuales que suplantan identidades a través de contenido hiperrealista y que podría representar una amenaza y un peligro para las personas ya que estos materiales son capaces de recrear acciones o hechos que nunca han sido llevados a cabo por humanos, lo cual socava la confianza en la evidencia audiovisual de manera radical (Chesney & Citron, 2019).

- Granjas de contenido y bots que se caracterizan por tener cuentas automatizadas organizadas en redes, que son impulsadas por IA y que comienzan a difundir información y narrativas falsas que crean conversaciones que generan percepción de apoyo popular espontáneo y de consenso social orgánico (Woolley & Howard, 2018).

Estas técnicas difuminan la línea entre lo real y lo creado, generando un clima de desconfianza generalizada, angustia social y una tendencia a la violencia simbólica (Bourdieu, 2000). Todo ello afecta directamente a la confianza sistémica, que el sociólogo Niklas Luhmann (1979) definió como el conjunto de mecanismos sociales basados en la expectativa de comportamientos predecibles y veraces, los cuales

permiten el funcionamiento eficiente de las instituciones y facilitan los vínculos sociales sin incurrir en excesivos costos de verificación.

Violencia digital catalizada por la IA: tipologías y casos

La violencia digital catalizada por la IA es un fenómeno multifacético que ha creado categorías inéditas de riesgos para las personas. Se caracteriza, además, por la dificultad que presenta para ser detectada (Chesney & Citron, 2019). La IA no es solo una amplificación de las antiguas formas de agresión en línea, sino que añade escalabilidad, personalización y adaptación de mensajes según las audiencias, así como una verosimilitud técnica sin precedentes.

Estas son algunas formas en que la inteligencia artificial se cristaliza y redefine el concepto de violencia digital.

- Ciberacoso (cyberbullying) amplificado por algoritmos.

Los contenidos falsos producidos por IA (deepfakes) provocan cambios radicales en las formas más básicas del ciberacoso. Las imágenes o audios alterados, así como los deepfakes creados con el objetivo de avergonzar o humillar a otras personas, se han convertido en instrumentos preferidos para el acoso digital, sobre todo entre los adolescentes.

73

El caso sucedido en la localidad de Almendralejo (España, 2023) es un ejemplo trágico de esta nueva realidad. Alumnos de un centro educativo de esta localidad extremeña convirtieron imágenes inocuas de la vida cotidiana de al menos veinte adolescentes en fotografías de desnudos mediante aplicaciones gratuitas de IA. Estas fotografías fueron luego compartidas con otros estudiantes, compañeros de clase de las víctimas, a través de grupos en WhatsApp y Telegram.

Los efectos psicológicos que esto provocó en las afectadas fueron desde un sentimiento de vergüenza paralizante, pasando por cuadros de ansiedad grave, hasta episodios de aislamiento social. Una agresión digital como esta afecta un elemento fundamental en el desarrollo de la identidad durante la adolescencia, una fase crucial para el crecimiento del yo. El hecho de que las víctimas no puedan interrumpir la propagación del contenido, junto a la naturaleza duradera de la huella digital, causa en ellas un intenso sentimiento de desesperanza. Esto es una expresión de lo que Seligman (1975) denomina “indefensión aprendida”, es decir, el estado de una persona que ha “aprendido” a actuar pasivamente, percibiendo que no tiene la facultad de intervenir.

Situaciones como esta ilustran la incapacidad cada vez mayor de la sociedad contemporánea para distinguir entre lo auténtico y lo artificial, lo que erosiona de manera sistemática la confianza en las demás personas (Luhmann, 1979).

- Violencia de género digital sistemática

La IA ha posibilitado la sistematización de la violencia digital de género por medio de deepfakes, que han permitido producir pornografía a gran escala.

Esta agresión perpetúa los patrones históricos de cosificación y subordinación, lo que Bourdieu (2000) definió como violencia simbólica: un proceso de dominación por medio de la distorsión de la realidad, a través de narrativas que destruyen la confianza social y naturalizan la desigualdad.

74

Un ejemplo de cómo la IA se utiliza para fomentar la violencia de género son plataformas como DeepNude (activa en 2019). El daño va más allá de lo digital: se han documentado casos de crisis de pánico, trastorno de estrés postraumático, desempleo e incluso reclusión social grave.

- Violencia política

Los deepfakes políticos son utilizados como armas de desestabilización geopolítica y polarización doméstica. El vídeo falso del presidente ucraniano Volodímir Zelenski llamando a la rendición de las fuerzas armadas de su país, que circuló en 2023 durante la guerra con Rusia, mostró cómo la IA puede minar la moral y sembrar confusión estratégica. A pesar de que el contenido fue desmentido con rapidez por el Digital Forensic Research Lab del Atlantic Council, alcanzó a más de 2 millones de usuarios antes de ser borrado, lo que evidenció la rapidez con que puede difundirse la desinformación política.

La mentira se transforma en una herramienta de poder cuando se institucionaliza, tal como Arendt (1969) previó. Hoy en día, esta institucionalización es algorítmica: para influir en la opinión pública a gran escala, tanto actores estatales como no estatales emplean microtargeting emocional, contenidos sintéticos y granjas de bots (Woolley & Howard, 2018).

Esta violencia política digital es una forma de “guerra cognitiva” en la que el campo de batalla es la percepción ciudadana, amenazando

el debate público informado.

- Violencia estructural simbólica y económica

Aunque a nivel simbólico la burla a través de memes producidos por IA pueda parecer una inocente manifestación de la cultura popular, con frecuencia esconde operaciones coordinadas que buscan desprestigiar y reforzar estereotipos.

En el ámbito económico, la falsificación digital puede dañar reputaciones profesionales construidas durante décadas. La habilidad de la inteligencia artificial para producir falsedades que parezcan creíbles puede ser usada para debilitar la credibilidad de individuos u organizaciones, lo cual conlleva perjuicios financieros debido a que se pierden contratos o por acoso en redes sociales.

La intersección entre la violencia económica y la simbólica muestra lo que Zuboff (2019) señala como una particularidad de la era digital: el poder de convertir la vivencia humana en datos y, posteriormente, esos datos en instrumentos para influir en la conducta. La IA no es un instrumento neutral, sino que fomenta las dinámicas de poder ya existentes.

75

La violencia digital provocada por la IA es el producto de una estructura técnico-operativa creada con la finalidad de optimizar la atención y el compromiso en las redes sociales. Hemos pasado de un modelo de desinformación artesanal a uno industrializado. La sinergia entre los algoritmos de las plataformas y sus usuarios es el elemento fundamental.

La arquitectura algorítmica de la viralización

La efectividad de la violencia digital catalizada por la IA es, en gran medida, un rasgo estructural del diseño de las plataformas digitales. La economía de la atención constituye el modelo de negocio que rige las redes sociales más grandes. Estas plataformas no están diseñadas para fomentar la verdad o el consenso, sino para maximizar el engagement y el tiempo que un usuario pasa frente a la pantalla. Esto da lugar a “cámaras de eco emocionales”: espacios digitales en los cuales la indignación y el miedo se transforman en divisas de visibilidad.

Aquí es donde el efecto de la inteligencia artificial se vuelve desbordante. Las noticias sensacionalistas o falsas son más difundidas y se

hacen virales con mayor rapidez que la información verdadera, llegando a más individuos en menos tiempo (Vosoughi, Roy y Aral, 2018). Esto ha sido demostrado por estudios del MIT. Los algoritmos de las plataformas, que tienen como objetivo recompensar el contenido que genera sorpresa, miedo o indignación, amplifican los efectos de la desinformación producida por IA.

Las investigaciones de Törnberg (2023) sobre la desinformación generativa han detectado tres rasgos técnicos que proporcionan a estos contenidos un potencial más elevado para volverse virales:

- Mejoramiento de las predicciones emocionales en donde los sistemas de inteligencia artificial tienen la capacidad de examinar distintas versiones de un contenido falso y seleccionar aquella que genera en públicos específicos los picos más elevados de reactividad emocional.

- Adaptación dinámica al contexto en la que los contenidos se modifican y se auto-adaptan en función de las circunstancias políticas, culturales o sociales locales, lo que incrementa su capacidad persuasiva en cada entorno.

- Sincronización estratégica que se relaciona la divulgación de contenidos con crisis auténticas o acontecimientos noticiosos para mejorar su credibilidad e influencia en el momento adecuado.

Las plataformas, además, monetizan la atención de las personas sin importar el contenido que consumen. En este sentido, la desinformación resulta más rentable que la información verificada, ya que las grandes plataformas capturan la mayor parte de los ingresos publicitarios digitales sin asumir los costos sociales que dicha desinformación produce.

Esta asimetría se agrava debido a la existencia de un verdadero “complejo industrial de la desinformación”: una red interdependiente compuesta por generadores de contenido, monetizadores de públicos, compañías de publicidad programática y empresas consultoras de microtargeting, que encuentran en la difusión viral de información falsa un modelo de negocio extraordinariamente rentable.

Vulnerabilidades psicológicas

El desarrollo de la desinformación impulsada por la inteligencia artificial no es únicamente resultado del progreso de las herramientas,

sino también de los cimientos frágiles de nuestras sociedades. Finalmente, la IA funciona como un espejo que muestra las debilidades de los individuos y de los sistemas.

Algunos de los principales elementos cognitivos que hacen a las personas más propensas a la desinformación producida por IA son:

- El sesgo de confirmación que hacen que los algoritmos vinculen a los usuarios con información que valida sus creencias ya existentes, fortaleciendo así sus prejuicios y esquivando la confrontación con datos que se opongan a ellas.

- La fluidez cognitiva que determina que los contenidos generados por IA, debido a su presentación cuidada y atractiva desde el punto de vista estético, suelen ser valorados más por su familiaridad y presentación que por su veracidad.

- El análisis crítico se vuelve complicado debido a la inmensa cantidad de datos disponibles. Cuando hay un exceso, solemos reaccionar desde lo emocional y compartir contenidos de manera impulsiva en vez de reflexionar sobre ellos.

77

Efectos estratégicos y psicosociales de la violencia digital intensificada por la IA

La IA trasciende la esfera individual y se transforma en un fenómeno con consecuencias psicosociales significativas en la estructura social. La conjunción de la verosimilitud, la permanencia y la escala algorítmica da lugar a un “trauma digital complejo”: un fenómeno patológico, de múltiples facetas, que impacta simultáneamente aspectos psicológicos, sociales y políticos de la experiencia humana.

Ahora examinaremos estos efectos a través del análisis individual, el social-comunitario y el estratégico-nacional.

- Consecuencias a nivel individual

La violencia digital se convierte en una patología psicológica compleja y persistente a nivel personal. Las víctimas atraviesan una serie de síntomas que reflejan la indefensión aprendida digital (Seligman, 1975): la sensación de que no se puede detener o controlar la propagación de los contenidos falsos lleva a una parálisis conductual, desesperanza y profundas alteraciones afectivas. Este fenómeno se presenta en varias formas específicas de afectación psicológica:

Síndrome de hipervigilancia digital: La vigilancia compulsiva en línea trae consigo trastornos de ansiedad generalizada, insomnio y fatiga mental.

Alteración de la identidad: Los ataques realizados a través de IA, sobre todo los deepfakes sexuales, interrumpen con violencia el proceso de construcción identitaria en los adolescentes. Esto ha sido llamado “fractura del ser digital” por los especialistas, es decir, una desconexión entre la identidad vivida y la identidad pública.

Estrés traumático: La persistencia de contenidos falsos y su repetición constante generan una condición crónica de victimización, similar a un trastorno de estrés postraumático. Esto provoca alteraciones en la percepción propia y en las relaciones emocionales con otros.

- Consecuencias a nivel social y comunitario

La IA fomenta la desinformación, lo que actúa como un catalizador de desintegración social y confirma las predicciones de Luhmann (1979) sobre la fragilidad de la confianza sistémica.

La IA puede utilizarse para atacar distintos niveles de confianza:

- Confianza interpersonal la cual instaura la duda sistemática sobre la veracidad de los contenidos y comunicaciones que las personas comparten entre sí.
- Confianza institucional que cuestiona la habilidad de los medios, las instituciones educativas y las autoridades para corroborar los hechos y transmitir la verdad.
- Confianza en los sistemas expertos donde se debilita la credibilidad de las técnicas y profesiones que se han considerado, por mucho tiempo, fuentes confiables de conocimiento.

Al mismo tiempo, se observa un proceso de aceptación paulatina de la violencia en formatos que son estéticamente atractivos, como los memes o los videos virales.

En la comunidad, además, se da el fenómeno de la fragmentación epistemológica: la coexistencia al mismo tiempo de realidades sociales que no se pueden reconciliar. Los grupos que están expuestos a diferentes entornos de desinformación producida por inteligencia artificial no solamente desarrollan perspectivas divergentes, sino también interpretaciones totalmente incompatibles entre sí, lo cual dificulta el diálogo.

go público y la negociación política. Consecuencias a nivel estratégico-nacional.

Desde el punto de vista estratégico, la desinformación generada por la inteligencia artificial constituye una amenaza transversal que pone en riesgo a organizaciones ya que es posible el sabotaje reputacional de empresas o sectores económicos estratégicos, dañando su credibilidad mediante campañas coordinadas de falsificaciones; e incluso la alteración de las percepciones de seguridad nacional a través de videos falsos de supuestos ataques o crímenes de guerra que debilitan la posición global de una nación y la moral de su ciudadanía.

Aspectos transversales, vulnerabilidades distintas

Para que el análisis de las consecuencias sea completo, es importante reconocer que el daño se distribuye de manera profundamente desigual entre diferentes grupos sociales. Las personas con discapacidad, las mujeres, las minorías étnicas y las comunidades LGBTQ+ a menudo sufren un impacto más fuerte, lo que agrava la discriminación estructural preexistente. Estos grupos vulnerables también tienen un acceso restringido a recursos técnicos, legales y de apoyo psicosocial para hacer frente a las agresiones. Por eso, tienen más posibilidades de ser blanco de campañas que fomentan el odio y la deshumanización. Esta faceta de la justicia social requiere que se tomen en cuenta intervenciones específicas y diferenciadas.

79

El panorama de la violencia digital promovida por inteligencia artificial es complicado y está claro que no existe una sola solución simple. Se necesita un marco estratégico que sea flexible y de múltiples niveles, el cual combine la prevención proactiva con mecanismos sólidos para contener y recuperar. Este planteamiento debe tener en cuenta que el desafío es tanto tecnológico como humano, tanto local como mundial, y requiere la cooperación de diversos actores sociales.

Una ciudadanía instruida y con capacidad de crítica es la primera línea de defensa frente a la desinformación. Es imprescindible una alfabetización digital crítica que promueva el pensamiento escéptico y la habilidad para discernir emocionalmente, lo que se cristaliza en medidas como:

- Transparencia tecnológica que instruye sobre el funcionamiento de las herramientas de IA generativa, así como sus capacidades y restricciones. Comprender que cualquier persona tiene la capacidad de crear un video falso hiperrealista con una aplicación gratuita en cuestión de

minutos es el primer paso hacia el desarrollo de un escepticismo saludable.

- Hábitos de verificación donde se busca promover el “pausar antes de compartir”. Fomentar la educación en comprobar las fuentes originales, recurrir a medios de confianza, emplear instrumentos para verificar datos y realizar búsquedas de imágenes inversas antes de divulgar información sospechosa.

- Inteligencia emocional digital que tiene como objetivo, capacitar a los usuarios para reconocer cuándo un contenido está diseñado para provocar indignación, miedo o risa. Esto permite crear un filtro emocional que prevenga la propagación impulsiva de falsedades.

- Programas educativos continuos que incorporan módulos de alfabetización mediática y digital en los planes de estudio. Mediante programas de educación continua, materiales accesibles y campañas públicas, es posible expandir estas iniciativas a las personas mayores y a los grupos vulnerables.

80

Se ha demostrado que la autorregulación de la industria no es suficiente. Es necesario contar con marcos legales firmes y coordinados que definan responsabilidades de manera precisa e implementen medidas disuasorias:

- Crear leyes que penalicen la creación y difusión de deepfakes y desinformación con intenciones maliciosas, especialmente en situaciones de desprestigio, acoso o extorsión, a partir de propuestas como la Ley de IA de la Unión Europea.

- Transparencia obligatoria que demande que los productores y las plataformas, indiquen de manera explícita el contenido generado por inteligencia artificial, notificando así al usuario mediante una señalización estandarizada y visible (como marcas de agua digitales, por ejemplo).

- Implementación de procedimientos legales eficientes para asegurar que las víctimas de perjuicios causados por estas prácticas reciban una compensación justa.

- Cooperación entre el sector público y el privado creando un ambiente de defensa en coordinación

La colaboración entre gobiernos, empresas tecnológicas, aca-

demia y la sociedad civil es crucial para crear un entorno de defensa más eficiente y dinámico que el de los autores de la desinformación.

Perspectiva de resiliencia estratégica

Ya que no es posible prevenir todos los ataques, se convierte en un objetivo estratégico el robustecimiento de las instituciones y la sociedad para soportar el impacto y evitar una caída sistémica ante la desinformación.

En este sentido, construir resiliencia implica la formación institucional que busca instaurar canales oficiales verificados, vías de comunicación nítidas y mensajes de respuesta ya preparados. Es esencial que los medios de comunicación, las organizaciones esenciales y las instituciones gubernamentales tengan la capacidad de reaccionar rápida y firmemente ante la desinformación virulenta.

También abarca el fortalecimiento de la confianza local al invertir en el capital social para reforzar la confianza dentro de las comunidades, disminuyendo su vulnerabilidad a la desinformación y ayudando a las víctimas a reintegrarse y obtener apoyo.

81

Y por último se quiere promover la diversificación del entorno informativo que se da cuando se respalda y se protege a los medios independientes y locales para diversificar las fuentes de información y minimizar la susceptibilidad de la sociedad a verse atrapada en burbujas de desinformación.

El desafío de la violencia digital catalizada por la IA y la desinformación ha dejado de ser una disputa tecnológica para convertirse en una crisis social. Nos encontramos en un cruce de caminos en el que el avance tecnológico pone en peligro la base más frágil y fundamental de la democracia: compartir una realidad común. La tarea más urgente no es detener la innovación, sino subordinarla a la dignidad humana mediante códigos éticos y legales, al mismo tiempo que formamos a la ciudadanía para que desarrolle la resiliencia crítica necesaria. El futuro de la paz social y la convivencia democrática en el siglo XXI dependerá de nuestra capacidad colectiva para proteger lo real y preservar los cimientos de la sociedad.

Referencias

Arendt, H. (1969). *On Violence*. Harcourt Brace & Company.

- Bourdieu, P. (2000). *Pascalian Meditations*. Stanford University Press.
- Chesney, R., & Citron, D. (2019). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*. Lawfare Research Paper Series.
- Luhmann, N. (1979). *Trust and Power*. John Wiley & Sons.
- Seligman, M. E. P. (1975). *Helplessness: On Depression, Development, and Death*. Freeman.
- Törnberg, P. (2023). How AI Is Changing the Nature of Disinformation. *Nature Human Behaviour*, 7(5), 654-665.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Council of Europe.
- Woolley, S. C., & Howard, P. N. (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press.
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.

La verdad bajo amenaza: desinformación y contenidos sintéticos en una nueva ecología informativa

83

Soledad Arréquez Manozzo

Emprendedora, periodista, profesora universitaria e investigadora especializada en desinformación y alfabetización mediática e informacional (AMli). doctoranda en comunicación (UCA) y especialista en educación y nuevas tecnologías (Facso, Argentina).

Actualmente asesora en comunicación en la defensoría del pueblo CABA. Cofundadora del proyecto Desconfío, una iniciativa para el combate contra la desinformación, y miembro del comité de organizador de la cumbre global sobre desinformación. Se desempeña como docente de grado y posgrado, coordina la agencia universitaria de noticias AUNO de la facultad de ciencias sociales de la Universidad Nacional de Lomas de Zamora (UNLZ) y está a cargo de la coordinación académica del departamento de CS de la comunicación de la Universidad Católica Argentina (UCA). Participa activamente en proyectos de formación para periodistas en América Latina. s.arreguez@gmail.com

Introducción

La inteligencia artificial generativa -un subcampo de la IA que permite crear contenidos originales en respuesta a *prompts*, es decir, instrucciones de un usuario- promueve profundas transformaciones en el esquema comunicacional del siglo XXI. Esta tecnología presenta nuevas sintaxis, dinámicas y especies mediáticas en una ecología informativa donde conviven las producciones humanas con contenidos sintéticos producidos por modelos de inteligencia artificial. Estas piezas artificiales pueden ser textos, imágenes, videos y audios capaces de reproducir de forma realista el estilo de escritura, la voz, la apariencia y hasta el movimiento de una persona que nunca existió. De igual modo, a partir de su entrenamiento, pueden retratar de forma sofisticada objetos y entornos, sin una referencia real,

La generación masiva de contenido sintético cambia el ecosistema comunicacional. Cada vez más contenido en Internet es generado por máquinas, a tal punto que el tráfico automatizado con el 51% superó al humano en 2025, de acuerdo al informe Bad Bots (Imperva, 2025). A su vez, la actividad de bots maliciosos representa el 37 % de todo el tráfico web, representando un aumento por sexto año consecutivo. Este incremento, advierten los investigadores de Imperva, se debe a la rápida adopción de la inteligencia artificial, en particular de los grandes modelos de lenguaje (LLM), que hacen más sencilla la creación de bots.

En ese sentido, la IA generativa modifica la forma en que las personas creamos, consumimos y comprendemos la información (Wang, 2023; Nielsen, 2025; López-Borrull & Lopezosa, 2025). En esta nueva ecología, estas piezas sintéticas problematizan la noción de “verdad” y de “evidencia” para la interpretación de la realidad. La noción de verdad ha sido ampliamente abordada por la filosofía desde Platón. Una corriente filosófica sostiene que las sociedades establecen en cada periodo histórico determinados criterios, colectivos y socialmente aceptados, para determinar la credibilidad de un hecho o fenómeno como así también para establecer los mecanismos para producir y difundir ese conocimiento (Kuhn, 1962; Foucault, 1975). Sin embargo, en la posverdad “los hechos objetivos son menos influyentes en la formación de la opinión pública que las apelaciones a la emoción y la creencia personal” (McIntyre, 2018).

La IA generativa plantea desafíos urgentes para la integridad informativa en un contexto de baja confianza en los medios, evasión noticiosa, fatiga informativa, técnicas de manipulación y narrativas desin-

formantes (Newman 2025).

Esta clase de contenidos pone a prueba las habilidades y competencias mediáticas e informacionales de la ciudadanía para distinguir un contenido real de uno creado artificialmente. El 58% de las personas encuestadas por el Digital News Report 2025 manifiesta inseguridad respecto de su capacidad para diferenciar lo verdadero de lo falso en noticias en línea. A su vez, una proporción considerable piensa que la IA provocará menos transparencia (-8), menos precisión (-8) y menos fiabilidad (-18) en las noticias (Newman, 2025). En sintonía, el Informe sobre el estado de la Década Digital 2025 señala que un 88% de los ciudadanos expresa inquietud por las *fake news* y la manipulación digital (Comisión Europea, 2024).

Los investigadores Cristian Vaccari y Andrew Chadwick (2020) realizaron un experimento para analizar hasta qué punto los *deepfakes* -una forma de desinformación audiovisual- engañan a las personas y demostraron que “las falsificaciones políticas profundas no necesariamente engañan a las personas, pero pueden sembrar incertidumbre que, a su vez, puede reducir la confianza en las noticias en las redes sociales” (párr. 51)

85

Por su parte, Barrington, Cooper, Farid (2025) realizaron un estudio para determinar cuánto están preparadas las personas para distinguir voces reales de otras generadas por IA, tanto en términos de coincidencia de identidad como de naturalidad. Sus resultados arrojaron que los participantes del estudio no pudieron identificar de forma consistente clones de voz generados por IA, por lo que no es fiable la percepción humana en estos casos (Barrington et al, 2025).

Los investigadores Chesney y Citron (2019) remarcan la capacidad de los *deepfakes* distorsionar la realidad y advierten que las “*fake news*” se extenderán a las “*deepfake news*” en el futuro:

A medida que las *deepfakes* se generalizan, el público puede tener dificultades para creer lo que ve o escucha, incluso cuando la información es real. A su vez, la propagación de *deepfakes* amenaza con erosionar la confianza necesaria para el funcionamiento eficaz de la democracia (2019, p.1786).

La inteligencia artificial generativa presenta tensiones y dilemas en torno a los sesgos en los modelos entrenados, la falta de transparencia en los sistemas, la privacidad y resguardo de datos de los usuarios y hasta la automatización de controles que podrían derivar en vigilancia.

Contenidos sintéticos para desinformar

El Informe de Riesgos Globales del Foro Económico Mundial (2025) señaló a la desinformación como uno de los riesgos más importantes para los próximos dos años. En ese documento, se remarca que “la desinformación y la información errónea también pueden ser el resultado de contenido manipulado por IA o errores humanos, y es probable que estos también aumenten ante el creciente volumen de contenido” (World Economic Forum, 2025, p.35).

La posibilidad de creación de piezas verosímiles, de forma sencilla a bajo costo, es utilizada por distintos actores para generar textos, imágenes, audio y hasta video de personas y hechos que jamás existieron, con el fin de producir piezas de desinformación, entendida como “todas las formas de falso, información inexacta o engañosa, diseñada, presentada y promovida para causar intencionalmente daño público o con fines de lucro” (Comisión Europea, 2018, p.5).

86 Para Wardle y Derakhshan (2017) la desinformación comprende información falsa, creada deliberadamente para causar daño, y los diferencia de los errores involuntarios (misinformation), y de la información basada en la realidad, utilizada para infligir daño a una persona, organización o país (malinformation). Por otra parte, los desórdenes informativos pueden asumir diferentes modalidades que incluye desde datos descontextualizados y contenido manipulado hasta parodia o sátira.

A la tipología desarrollada por Wardle (2019), la inteligencia artificial potencia el contenido fabricado, entendido como “un contenido nuevo 100% falso, diseñado para engañar y hacer daño” (p.51), a partir de la creación de *deepfakes*. Estas falsificaciones profundas consisten en videos generados por algún tipo de inteligencia artificial para representar de forma realista a una persona haciendo o diciendo algo que nunca hizo. En Internet circularon *deepfakes* de personas públicas como Donald Trump, Barack Obama, Taylor Swift, Lula Da Silva y Elon Musk, entre otros.

A diferencia del contenido manipulado o fabricado (Wardle, 2019), estas creaciones no requieren de un material original como punto de partida para la generación del archivo. Estos modelos de IA pueden crear una imagen de una persona pública a partir de un *prompt*.

Los investigadores Alexandre López-Borrull y Carlos Lópezosa (2025) advierten que la IA generativa también facilita la difusión de desinformación al hacerla más específica, personalizada y escalable.

La combinación de contenido sintético y algoritmos de recomendación de plataformas en línea amplifica el alcance de las narrativas falsas, a menudo fuera del control de los mecanismos de supervisión tradicionales (p.14).

El experto en IA Lukas Görög aseguró en la Cumbre Global sobre Desinformación 2024 que la IA generativa esparció su influencia en desinformación en distintos ámbitos: aumentó la pornografía con *deep-fakes* en más de 464 %; las estafas por email mediante ‘phishing’ crecieron 1.256 %; los ciberataques aumentaron en un 104%, y los sitios de desinformación creados con inteligencia artificial se incrementaron un 1.124 %. Chesney y Citron (2019) advierten sobre los riesgos que los *deepfakes* representan para la democracia:

Las falsificaciones profundas permitirán a las personas vivir en sus propias realidades subjetivas, donde las creencias pueden sustentarse en “hechos” inventados. Cuando las ideas empíricas básicas provocan una acalorada controversia, el discurso democrático tiene dificultades para avanzar. En un mercado de ideas inundado de vídeos y audios falsificados, los hechos veraces tendrán dificultades para emerger de la confusión. (p. 1778)

87

Sin duda, la desinformación tiene efectos dañinos para el funcionamiento de las sociedades democráticas; y los contenidos sintéticos acentúan los riesgos de ser víctimas de mentiras y engaños al poner en jaque los marcos tradicionales de evidencia y comprobación. Ya no alcanza con “ver para creer”, porque la inteligencia artificial produce verosimilitud y hasta logra engañar a los sentidos.

Herramientas digitales para la verificación

La tecnología digital es una aliada a la hora de contrarrestar la desinformación. Los periodistas y fact checkers alrededor del mundo complementan su práctica profesional con herramientas digitales para la verificación de fechas, dominios web, propiedad de empresas, perfiles en línea, cuentas automatizadas en redes sociales, imágenes, videos, geolocalización y hasta el estado del clima en un día y fecha particular, entre otros datos.

Los buscadores inversos de imágenes como Google Images, Tin-Eye y Yandex, permiten identificar si una imagen se encuentra publicada en Internet lo que brinda indicios sobre su procedencia. No son capaces de identificar alteraciones, recortes u otras ediciones de la imagen. Por ese motivo, estas aplicaciones permiten identificar rápidamente

piezas desinformantes que buscan descontextualizar hechos o situaciones: imágenes viejas son atribuidas a hechos recientes, generando confusión en la audiencia sobre la referencia temporal, y muchas otras, sin corresponder tampoco, con la ubicación.

Los indicios visuales de las imágenes se pueden verificar con las aplicaciones de Google Maps, Google Street View y Google Earth que permiten confirmar ubicaciones de imágenes o vídeos a partir de la identificación de paisajes, calles, monumentos y otros sitios de interés. Estas herramientas pueden complementarse con publicaciones en redes sociales que están georreferenciadas y/o último momento que brinden mayor información del espacio a verificar.

Para la verificación de videos en YouTube, la extensión InVID es una de las opciones más completas. Emplea una búsqueda inversa de los fotogramas clave del video para identificar publicaciones anteriores. Esta herramienta no da cuenta de manipulaciones y/o ediciones del video original. También brinda una opción de lupa para analizar imágenes y otra para visualizar los metadatos de imágenes (jpeg) y vídeos (en formato m4v mp4), entre otras funciones.

La verificación en la era de la inteligencia artificial

Los medios de comunicación incorporan la inteligencia artificial para la búsqueda, creación y distribución de contenidos. La IA tiene potencial para robustecer el chequeo de información, el procesamiento y análisis de grandes caudales de datos, la detección de patrones, el monitoreo de tendencias, la edición de textos, y la identificación de sesgos e inconsistencias, entre otras (Montoro-Montarrosó, 2023; Pino, A. & Arréquez Manozzo, 2025).

Los medios de comunicación pueden utilizar la IA generativa para combatir la desinformación. Existen distintos programas en línea que permiten comprobar contenidos sintéticos de forma sencilla y rápida. Sin embargo, hay que remarcar que ninguna herramienta es infalible, por lo que se sugiere como complemento de la investigación periodística, no como su sustituto.

. GPTZero es un software online para detectar texto generado por la mayoría de los modelos de IA. Está entrenado con una colección grande y variada de textos escritos tanto por humanos como por IA. Puede proporcionar evaluaciones reales sobre si un documento fue generado por IA en función de su proceso de escritura.

. Hive Moderation es una plataforma basada en inteligencia artificial que detecta, clasifica y modera contenidos en tiempo real. Utiliza modelos entrenados para identificar imágenes, videos y textos creados por inteligencia artificial.

. SyntheticEye es una plataforma que combina dos aplicaciones para la verificación de contenido: Aletheia diferencia los rostros humanos reales de los generados por IA mediante técnicas avanzadas de aprendizaje automático, y Argus, que identifica varios tipos de imágenes generadas por IA mediante aprendizaje automático de vanguardia.

Con la popularización de los modelos de lenguaje grandes (LLM) muchos usuarios preguntan a chatbots por la veracidad de las imágenes. Un estudio de Centro Tow para el Periodismo Digital encontró que los siete modelos estudiados no lograron identificar correctamente la procedencia de las fotos de forma consistente. Sin embargo, cuando las herramientas tradicionales no funcionan, los LLM podrían ser útiles para identificar la procedencia de las imágenes, ya que ofrecen puntos de partida útiles para los fact checkers (Chandrasekar & Jazwińska, 2025).

89

A pesar del avance tecnológico, aún no hay una forma perfecta para detectar si un contenido fue generado por IA. Y, a partir de la sofisticación de las tecnologías, cada vez será más difícil determinar la veracidad de un contenido basándonos exclusivamente en ese material. Frente a las formas sintéticas de desinformación, es importante combinar las herramientas digitales con las prácticas y técnicas del periodismo tradicional: buscar diversidad de fuentes, contrastar información, escuchar voces alternativas, y contextualizar datos, entre otras.

La ciudadanía, por su parte, ensaya diferentes estrategias para la verificación de los contenidos en línea: la contrastación de noticias con otros medios y fuentes alternativas, consultas a fuentes expertas o especializadas, y hasta preguntas en motores de búsqueda y chatbots. A estas prácticas, se suman otras estrategias específicas frente a los contenidos sintéticos: verificar si los rostros presentan simetrías perfectas y los cuerpos algunas deformidades, si los fondos están muy brillantes, desenfocados o muestran patrones repetitivos, si las fuentes de luz no coinciden con la iluminación de los objetos y no se alinean con sus respectivas sombras, y si no hay sincronización entre sonido e imagen (sincronización de labios), entre otras.

Dado que la tecnología avanza, muchos de estos errores se corregirán en breve, por lo que es vital prestar atención a la información vi-

sual, desde las letras en carteles hasta el paisaje o estilo arquitectónico, que permita dar contexto a lo que se muestra. En esta nueva ecología informativa, se requiere de una ciudadanía alfabetizada mediáticamente, con una mirada atenta y una actitud escéptica, que permita detenerse, analizar y reflexionar sobre lo que se observa para determinar su grado de credibilidad.

Conclusiones

Los contenidos sintéticos ponen en tensión las formas tradicionales para establecer verosimilitud y credibilidad de textos, imágenes y videos. Ya no basta con “ver” para creer, dado que la inteligencia artificial logró reproducir sujetos, objetos y entornos de forma hiperrealista. Esta tecnología aplicada en campañas de desinformación aumenta los riesgos de ser víctimas de engaños y mentiras. Los contenidos sintéticos desinformantes, lejos de ser piezas aisladas, forman parte de una nueva modalidad de causar daño o perjuicio a una persona, empresas o colectivo.

Frente a este escenario, la Nueva Teoría Estratégica (NTE) brinda un marco de pensamiento estratégico, complejo, relacional y más humano, para pensar la comunicación ante las narrativas desinformantes. El modelo sociocomunicacional propone la co-construcción, co-creación y co-evolución, entendiendo que la comunicación no es ‘dada’ sino que se construye en una actividad coordinada con otros actores. Herrera Echenique y Untiveros (2020) señalan que

analizar la comunicación no es sólo asunto de tecnologías de información o dispositivos, sino de relaciones entre personas, grupos, organizaciones y la sociedad misma. La comunicación ha cambiado el modo de convivencia de los seres humanos y el modo como se relacionan los actores de la comunicación (p.113).

En tiempos de contenidos sintéticos, el desafío de la comunicación parece ser encontrar dinámicas y prácticas que habiliten conversar, debatir y generar consensos sobre los asuntos públicos en sociedades cada vez más polarizadas.

En esta nueva ecología se requiere que las soluciones contra la desinformación sean estratégicas y multidimensionales desde los gobiernos locales y nacionales, las grandes empresas tecnológicas, los medios de comunicación, y la sociedad civil. Las acciones deberían con-

trarrestar las piezas desinformantes, pero también construir confianza con las audiencias que habilite el debate y diálogo ciudadano. Para esa tarea, la alfabetización mediática e informacional es vital para que la ciudadanía pueda participar de forma activa en su comunidad. Las personas que pueden acceder, evaluar, analizar y crear mensajes mediáticos tendrán más herramientas para ejercer su ciudadanía en el siglo XXI.

Desde la industria de los medios, el combate de la desinformación requiere de inversión y capacitación en herramientas digitales de verificación como así también de metodologías que recuperen lo mejor del oficio: hacer preguntas, buscar fuentes, indagar en lo no dicho y desconfiar del poder. Ejercer el periodismo en la era de los contenidos sintéticos requiere reforzar los principios por la búsqueda de la verdad, la rigurosidad de la información y la responsabilidad en la publicación.

Referencias

- Barrington, S., Cooper, EA y Farid, H. (2025). Las personas están mal equipadas para detectar clones de voz generados por IA. *Nature* . 91
<https://doi.org/10.1038/s41598-025-94170-3>
- Comisión Europea. (2024). *State of the Digital Decade 2025 report. Dirección General de Redes de Comunicación, Contenido y Tecnologías*.
<https://digital-strategy.ec.europa.eu/es/library/state-digital-decade-2025-report>
- Chandrasekar, A., & Jaźwińska, K. (2025, agosto 26). Why AI models are bad at verifying photos. *Columbia Journalism Review*. https://www.cjr.org/tow_center/why-ai-models-are-bad-at-verifying-photos.php
- Chesney, R., & Citron, D. K. (2019). Deep fakes: Un desafío inminente para la privacidad, la democracia y la seguridad nacional. *California Law Review*, 107, 1753–1819.
- Echenique, R. H., & Untiveros, R. B. (2020). De La Comunicación Instrumental a la Estratégica. *A Visão de 23 Renomados Autores em 5 Países*, 108.
- Foucault, M. (1975). *Vigilar y castigar: El nacimiento de la prisión* (A. Garzón del Camino, Trad.). Siglo XXI.
- Guevara-Chávez, J. & Almeida-Macias, M. (2025). Infodemia y alfabetización mediática frente a los riesgos de la desinformación automatizada en entornos digitales.. *Scripta Mundi*, 4(2), 10-36.
<https://doi.org/10.53591/scmu.v4i2.2309>
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
- López-Borrull, A. & Lopezosa, C. (2025). Mapeo del impacto de la IA gen-

erativa en la desinformación: Perspectivas de una revisión exploratoria. *Publicaciones*, 13 (3), 33. <https://doi.org/10.3390/publications13030033>

McIntyre, L. (2018). *Post-truth*. MIT Press.

Montoro-Montarroso, Andrés; Cantón-Correa, Javier; Rosso, Paolo; Chulvi, Berta; Panizo-Lledot, Ángel; HuertasTato, Javier; Calvo-Figueras, Blanca; Rementeria, M. José; Gómez-Romero, Juan (2023). "Fighting disinformation with artificial intelligence: fundamentals, advances and challenges". *Profesional de la información*, v. 32, n. 3, e320322. <https://doi.org/10.3145/epi.2023.may.22>

Newman, N. (2025, 17 de junio). *Digital News Report 2025: Resumen ejecutivo*. Reuters Institute. <https://reutersinstitute.politics.ox.ac.uk/es/digital-news-report/2025/dnr-resumen-ejecutivo>

Pino, A., & Arréguez Manozzo, S. (2025). The trust editor: An AI tool from proyecto Desconfío to recover the trust in news. In H. M. Sánchez Gonzales (Ed.), *Entrepreneurship and Artificial Intelligence in Journalism* (pp. 95–112). McGraw-Hill.

Imperva. (2025). *2025 bad bot report*. <https://www.imperva.com/resources/resource-library/reports/2025-bad-bot-report/>

Vaccari, C. & Chadwick, A. (2020). Deepfakes y desinformación: Explorando el impacto del video político sintético en el engaño, la incertidumbre y la confianza en las noticias. *Social Media + Society*, 6 (1). <https://doi.org/10.1177/2056305120903408>

Wang, Y. (2023, agosto 18). *Realidades sintéticas en la era digital: Explorando las oportunidades y los desafíos del contenido generado por IA* [Preprint]. *TechRxiv*. <https://doi.org/10.36227/techrxiv.23968311.v1>

Wardle, C. & Derakshan, H. (2017). *Information Disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe.

Wardle, C. (2019). *First draft's essential guide to understanding information disorder*. First Draft.

La desinformación se ha convertido en uno de los grandes desafíos de nuestro tiempo. La información fluye sin control y la verdad compite con narrativas que están diseñadas para influir y/o manipular. Este libro aborda, desde una mirada crítica y multidisciplinaria, cómo los entornos digitales han transformado la manera en que construimos credibilidad, consumimos contenidos y tomamos decisiones en la vida cotidiana, política y organizacional.

Con las aportaciones de especialistas en comunicación, estrategia, tecnología y comportamiento humano, la obra explora el origen de la desinformación: desde los sesgos cognitivos hasta las dinámicas de las redes sociales y el impacto creciente de la inteligencia artificial en la producción de contenidos. El lector encontrará aquí no solo un diagnóstico, sino una comprensión profunda de un fenómeno complejo y en constante evolución.

El libro sitúa la gestión de la desinformación como un desafío estratégico que exige nuevas formas de pensar y actuar. Se propone abandonar modelos tradicionales centrados en el control del mensaje para avanzar hacia una comunicación basada en la construcción de sentido, la interacción y la confianza. En este contexto, la credibilidad emerge como un activo clave que debe ser gestionado con visión, ética y responsabilidad.

De igual manera, la obra incorpora una perspectiva prospectiva que invita a mirar más allá del presente. ¿Cómo evolucionarán los ecosistemas informativos? ¿Qué papel jugarán las tecnologías emergentes en la configuración de la verdad? ¿Estamos preparados para enfrentar los riesgos y aprovechar las oportunidades?

Dirigido a académicos, profesionales, líderes organizacionales y a todo aquel interesado en comprender el impacto de la desinformación en nuestra sociedad actual, este libro nos invita a reflexionar, cuestionar y actuar.

Octavio Islas
Haydée Guzmán
Antonio Salgado
Juan Carlos Peña
María Eugenia Garcés
Maritza Mosquera de Sumich

Amaia Arribas
Dona Arrieta
Alejandro Salum
Gabriel Salazar
Jorge Rendón
Soledad Arréquez

ISBN: 978-9907-0-1255-2

