La IA, negociación y el nuevo arte de moldear opiniones

Cedric Schweizer

Chrabieh & Schweizer Consultancy

https://doi.org/10.26439/puentes.comunicacion2026.8872

RESUMEN. Esta investigación parte de una constatación que, desde mi experiencia en negociación y análisis de riesgo, resulta cada vez más difícil de ignorar: la recolección masiva de datos digitales —capaz de perfilar comportamientos, emociones y preferencias con una precisión milimétrica— ha sentado las bases de una nueva forma de persuasión más silenciosa y efectiva que cualquier discurso clásico. Con la llegada de la inteligencia artificial (IA), que procesa esa información a gran escala y en tiempo real, el arte de influir en la opinión pública ya no se construye solo con palabras, sino con arquitectura algorítmica. Desde la retórica de la antigüedad hasta los sistemas de recomendación actuales, este artículo analiza cómo han evolucionado las estrategias de influencia y cuáles son los riesgos que enfrentamos en esta nueva era de persuasión automatizada, íntima y personalizada.

INTRODUCCIÓN

Desde tiempos remotos, influir en lo que otros piensan ha sido una habilidad esencial para la vida colectiva. En la Atenas del siglo IV a. C., los sofistas y oradores como Demóstenes entendían que quien dominaba la palabra dominaba el poder. Aquella lógica de influencia, basada en el uso estratégico del lenguaje, sigue vigente. Lo que ha cambiado es el canal, la velocidad y, sobre todo, el grado de conciencia del receptor. Hoy, la persuasión ya no se basa únicamente en la retórica humana, sino en sistemas automatizados que recolectan datos, predicen comportamientos y adaptan mensajes de manera casi instantánea.

El auge de la inteligencia artificial (IA) ha configurado un nuevo paradigma persuasivo más preciso, más veloz y, lo que me resulta más inquietante, profundamente imperceptible. Mientras que la persuasión clásica apelaba a la razón, a la emoción y a la credibilidad del emisor (Aristóteles, ca. 350 a. C./1999), el ecosistema digital actual se sostiene sobre algoritmos que analizan

sentimientos, detectan patrones invisibles y personalizan contenidos sin necesidad de activar nuestras defensas racionales.

Esta transformación no es exclusiva del ámbito comercial o del entretenimiento, sino que ha comenzado a filtrarse en espacios como la política, la justicia o la diplomacia, donde herramientas algorítmicas potencian métodos tradicionales de negociación que permiten moldear el comportamiento humano con eficacia inédita. En este contexto, Zuboff (2019) advierte, con razón, sobre el surgimiento de una forma avanzada de "capitalismo de la vigilancia" en el cual las decisiones no solo se registran, sino que se predicen e incluso se reconfiguran antes de que el individuo las tome.

Frente a este panorama, este artículo propone una revisión crítica de cómo la combinación entre la recolección masiva de datos digitales, la inteligencia artificial y los principios propios de la negociación conductual está redefiniendo los mecanismos mediante los cuales se construyen, manipulan y reafirman las opiniones públicas. La propuesta se articula desde una mirada interdisciplinaria que combina nociones de filosofía política, psicología cognitiva y comunicación estratégica.

La estructura del artículo se organiza en cuatro secciones. En primer lugar, se repasan los hitos históricos de la persuasión desde la retórica clásica hasta el giro audiovisual de la era televisiva. Luego, se describe el surgimiento de un nuevo ecosistema digital marcado por la omnipresencia de datos, la lógica algorítmica y la personalización automatizada de mensajes. En tercer lugar, se analiza cómo modelos diseñados para contextos de negociación extrema, como el modelo escalonado del FBI, están siendo trasladados al ámbito digital. Finalmente, se discuten las implicancias éticas y sociales de estas dinámicas, así como los posibles caminos para fortalecer la capacidad crítica y deliberativa de los ciudadanos frente a formas de influencia cada vez más invisibles y eficaces.

BREVE HISTORIA DE LA PERSUASIÓN

La persuasión ha acompañado al ser humano desde que comenzamos a organizarnos políticamente. No es una invención moderna, sino una práctica que ha evolucionado junto con los medios de comunicación, la tecnología y las estructuras de poder. Comprender su recorrido histórico no es un ejercicio académico vacío; al contrario, es clave para dimensionar los cambios radicales que hoy introduce la inteligencia artificial en el modo en que se moldea la opinión pública.

Retórica clásica

En la antigua Grecia, la persuasión era una herramienta fundamental para la vida cívica. Aristóteles, en la *Retórica*, identificó tres pilares que siguen vigentes incluso en campañas digitales del siglo XXI: el *ethos*, la credibilidad del emisor; el *pathos*, la apelación emocional, y el *logos*, la fuerza lógica del argumento.

Oradores como Demóstenes en Grecia o Cicerón en Roma fueron maestros del arte de convencer con la palabra. Sus discursos no solo influían en decisiones judiciales o políticas, sino también moldeaban el imaginario colectivo. En aquella época, persuadir constituía una práctica visible y

pública, estrechamente vinculada a la presencia del orador, su gestualidad, su voz y su habilidad para suscitar emociones a través de recursos retóricos.

La retórica no era un accesorio, era una disciplina formal y parte esencial de la formación ciudadana. Como señala Kennedy (2007): sentó las bases para comprender el lenguaje no solo como medio de comunicación, sino como herramienta de poder, de influencia y de deliberación política.

Revoluciones mediáticas

El siguiente gran salto en esta historia llegó con la imprenta de Gutenberg en el siglo XV. Gracias a ese avance, la palabra impresa se volvió reproducible, portable, multiplicable y, con ello, se amplificó su capacidad persuasiva. Panfletos como los de Martín Lutero no solo impulsaron debates teológicos, también reconfiguraron el poder político y religioso mostrando cómo una tecnología podía catalizar una transformación social a escala.

Durante los siglos XVIII y XIX, la prensa escrita se consolidó como el principal espacio donde se moldeaba la opinión pública. La prensa partidista, primero, y la prensa amarilla, después, mostraron que no todo pasaba por el argumento racional: la emoción, el escándalo y la repetición empezaban a formar parte de la caja de herramientas persuasiva.

El siglo XX trajo consigo nuevos lenguajes: la radio, el cine y la televisión. Durante la Segunda Guerra Mundial, Goebbels supo utilizar la radio como un canal directo de propaganda. En paralelo, Estados Unidos recurría a directores como Frank Capra para producir documentales que justificaran su intervención bélica. Se adaptaba la lógica técnica y narrativa de cada medio para convertirlo en herramienta persuasiva y, con ello, se expandía el campo de la persuasión (Jowett & O'Donnell, 2018).

La era televisiva y el poder de la imagen

La televisión cambió las reglas del juego. Incorporó algo que los medios anteriores apenas sugerían: la imagen en movimiento, la estética visual como parte decisiva del mensaje. Ya no bastaba con tener argumentos sólidos o un tono emocional eficaz, ahora importaba la postura, la mirada, la apariencia del emisor.

El debate presidencial entre Richard Nixon y John F. Kennedy en 1960 es un hito paradigmático. Quienes lo escucharon por radio consideraron que Nixon había ganado. Quienes lo vieron por televisión, en cambio, se inclinaron por Kennedy: más joven, más carismático, más telegénico. Nixon, por el contrario, aparecía pálido, sudoroso, agotado. Esa diferencia de percepción marcó el inicio de la videopolítica, donde la lógica y la estética empezaron a tener el mismo peso en la construcción del liderazgo público (Jamieson & Campbell, 2001; McLuhan, 1964).

La televisión permitió también campañas publicitarias más sofisticadas, basadas en la repetición, la segmentación por horarios y la carga emocional de las imágenes. Como señala Postman (1985): no solo cambió lo que comunicábamos, también transformó cómo pensábamos, cómo recordábamos y cómo debatíamos en sociedad.

De la persuasión visible a la influencia encubierta

Hasta finales del siglo XX, la mayoría de los actos persuasivos eran detectables: discursos, anuncios, debates públicos. El receptor sabía que alguien estaba intentando convencerlo. Con la revolución digital, esa claridad se diluye. Hoy, la persuasión se integra en la rutina, se esconde detrás de una notificación, de una sugerencia personalizada, de un *scroll* infinito.

Como advierte Sunstein (2017), ya no es necesario imponer un punto de vista ni argumentar frontalmente, basta con diseñar arquitecturas de información que empujen decisiones sin que el usuario lo note. Esa transición de lo visible a lo invisible es uno de los cambios más profundos que subyace a este análisis.

EL NUEVO PARADIGMA: DATOS, INTELIGENCIA ARTIFICIAL Y LA ARQUITECTURA DE LA PERSUASIÓN DIGITAL

El auge de la persuasión digital no comenzó directamente con la inteligencia artificial, lo hizo antes con la capacidad de recolectar datos digitales de forma masiva, detallada y a nivel individual. Es esa acumulación silenciosa de información —sobre emociones, rutinas, hábitos de consumo o patrones psicológicos— la que ha sentado las bases para una nueva forma de influencia más precisa, más estratégica y, sobre todo, más difícil de detectar.

Luego, con la llegada de la inteligencia artificial, ese caudal de datos pudo ser procesado, interpretado y activado a una escala nunca vista. Lo que antes requería intuición y tiempo humano, hoy se ejecuta en milisegundos. Las estrategias tradicionales de persuasión —centradas en discursos, carisma y visibilidad— han sido reemplazadas por algoritmos que influyen sin rostro, sin voz y sin pedir permiso.

Hoy, la influencia ya no necesita de un orador brillante ni de una campaña masiva; basta con un algoritmo bien entrenado. Las plataformas digitales no solo recopilan y analizan datos, sino que también generan interacciones persuasivas altamente personalizadas, muchas veces sin intervención humana directa. Un ejemplo de ello puede observarse en TikTok, cuyo sistema de recomendación ha sido cuestionado por su capacidad para intensificar determinados estados emocionales. Investigaciones periodísticas mostraron que bastaban pocas interacciones con contenidos relacionados con tristeza, ansiedad o aislamiento para que el algoritmo comenzara a recomendar de manera recurrente videos vinculados con depresión, autolesión o discursos extremistas, lo que generaba una espiral de refuerzo emocional difícil de percibir por el usuario.

En estos casos, la persuasión ya no opera mediante discursos públicos dirigidos a una audiencia colectiva, sino a través de flujos de contenido personalizados que moldean progresivamente la percepción, las emociones y el comportamiento de cada individuo. Casos como el de Cambridge Analytica, que se desarrollará más adelante, evidencian hasta qué punto estos mecanismos pueden ser utilizados también con fines políticos y electorales.

Este nuevo paradigma descansa sobre tres pilares tecnológicos: la conectividad permanente; el procesamiento masivo de datos, también conocido como *big data*; y los algoritmos de recomendación

que actúan como mediadores invisibles entre lo que queremos, lo que creemos querer y lo que finalmente recibimos (Eubanks, 2018; Zuboff, 2019).

Big data y *microtargeting*

Cada clic, búsqueda, desplazamiento o "me gusta" deja una huella. Esas huellas, que parecen inofensivas, permiten construir perfiles psicográficos de una profundidad insospechada. Sistemas de inteligencia artificial entrenados con millones de registros son capaces de describir quiénes somos y de anticipar cómo podríamos reaccionar ante un estímulo futuro. Zuboff (2019) ha denominado a esta lógica de extracción *excedente conductual*: el corazón de la nueva persuasión digital.

Este conocimiento se traduce en una microsegmentación comunicativa en la que ya no se busca transmitir un mismo mensaje a toda la población, sino elaborar versiones diferenciadas adaptadas a segmentos específicos según sus intereses, valores, temores o estados emocionales. En este contexto, una misma campaña puede enfatizar aspectos distintos —e incluso aparentemente incompatibles— dependiendo del perfil del receptor. Así, por ejemplo, un candidato político puede presentarse ante un grupo como defensor del crecimiento económico y, ante otro, destacar prioritariamente su compromiso con la justicia social o la protección ambiental. Aunque el objetivo general de la campaña permanece intacto, el discurso se ajusta estratégicamente para conectar con las sensibilidades particulares de cada audiencia.

Este tipo de segmentación no solo redefine las estrategias de comunicación, sino que también transforma las reglas del juego democrático. La personalización algorítmica permite adaptar mensajes en tiempo real para reforzar creencias previas, activar determinadas emociones o incentivar comportamientos específicos sin que el receptor sea plenamente consciente de los mecanismos de personalización que operan detrás de la comunicación. De este modo, la persuasión deja de desarrollarse en un espacio público compartido y verificable para desplazarse hacia experiencias comunicativas individualizadas y difíciles de contrastar colectivamente.

Personalización y sesgos cognitivos

Los algoritmos de recomendación están presentes en casi todos los espacios digitales que usamos a diario: desde YouTube hasta TikTok, desde Netflix hasta Spotify. Estos sistemas no se limitan a ofrecer contenido que podría interesarnos, sino que buscan maximizar nuestro tiempo de permanencia, nuestra atención, nuestra dependencia. Y en ese proceso, no solo responden a nuestros gustos: también los moldean.

Esta personalización no es neutra. Al ofrecernos constantemente contenido alineado con nuestras creencias previas, los algoritmos refuerzan lo que ya pensamos, validan lo que ya sentimos y dificultan el pensamiento crítico. Este fenómeno, conocido como sesgo de confirmación, ha sido ampliamente estudiado por la psicología social (Kahneman, 2011) y hoy es una de las bases funcionales de la arquitectura digital.

El resultado son cámaras de eco: entornos cerrados donde solo circula lo que coincide con nuestra visión del mundo. Allí se reduce el disenso, se diluye la duda y el diálogo plural se vuelve

cada vez más improbable. Pero el sesgo de confirmación no es el único recurso que explota la IA, también entran en juego otros mecanismos más sutiles, pero igual de eficaces:

- El efecto halo, que lleva a atribuir cualidades positivas a un mensaje solo porque proviene de una fuente admirada.
- El sesgo de anclaje, que utiliza datos o imágenes impactantes como punto de referencia para influir en nuestras valoraciones posteriores (Tversky & Kahneman, 1974).
- La reciprocidad simbólica, que se activa cuando una marca o figura pública nos "ofrece" algo gratuitamente, generando en nosotros un sentido de deuda emocional (Cialdini, 2007).

Lo más preocupante es que todo esto ocurre sin que el receptor lo perciba, pues la persuasión ya no se impone, se filtra. Se vuelve transparente, casi imperceptible. Opera en segundo plano, como si fuese parte del flujo natural de nuestra rutina digital. Como bien señala Eubanks (2018), los sistemas algorítmicos no solo observan lo que hacemos, también reconstruyen quiénes somos, y nos redibujan como sujetos de consumo y como objetos de influencia política.

EL CASO CAMBRIDGE ANALYTICA: PSICOGRAFÍA, DATOS Y ELECCIONES

A lo largo de los últimos años, he podido observar, con creciente interés profesional, cómo este nuevo ecosistema digital ha transformado silenciosamente la forma en que se configuran las decisiones colectivas. Lo que antes requería un discurso, una presencia o una negociación explícita, hoy puede activarse desde un simple algoritmo invisible, pero altamente eficaz.

En este sentido, hay un caso que sintetiza con particular claridad esta convergencia entre datos, psicología y poder político: Cambridge Analytica. Más que un episodio aislado, lo considero un punto de inflexión que nos obliga a repensar cómo entendemos la influencia y qué lugar ocupa la voluntad individual en entornos dominados por la inteligencia artificial.

Origen y metodología de trabajo

Cambridge Analytica, activa entre 2013 y 2018, fue una filial de la empresa británica SCL Group y contó con financiamiento ligado al entorno del Partido Republicano de Estados Unidos. Su propuesta combinaba tres ingredientes altamente eficaces: la explotación de macrodatos (*big data*), la aplicación del modelo psicográfico Openess Conscientiousness Extraversion Agreeableness Neuroticism (OCEAN) y técnicas de microsegmentación algorítmica para intervenir en procesos electorales.

Uno de sus métodos más polémicos consistió en recolectar datos personales a través de una simple aplicación de prueba de personalidad en Facebook. Lo grave no fue solo el acceso directo al perfil del usuario, sino la recolección no autorizada de datos de sus contactos. Con esta base, se generaron perfiles psicológicos detallados y se entrenaron algoritmos para enviar mensajes personalizados diseñados para activar emociones precisas y modificar comportamientos (Kosinski et al., 2013).

Aplicación del modelo OCEAN

El modelo OCEAN —que analiza apertura, responsabilidad, extraversión, amabilidad y neuroticismo— fue central para esta estrategia. Permite clasificar a las personas según sus rasgos de personalidad y adaptar los mensajes a su perfil emocional. Por ejemplo, quienes mostraban altos niveles de neuroticismo eran bombardeados con mensajes que apelaban a amenazas o inseguridad; mientras que los perfiles responsables recibían contenidos centrados en el orden, la legalidad o la estabilidad.

La IA hacía posible aplicar este enfoque a gran escala midiendo resultados en tiempo real y ajustando las campañas en función de la respuesta emocional de cada grupo (Matz et al., 2017). Lo significativo de este proceso es que una herramienta pensada para describir perfiles humanos fue convertida en una estrategia de intervención política altamente segmentada y automatizada sin siquiera requerir interacción directa.

Principales escenarios de intervención

Aunque el caso más conocido fue la campaña presidencial de Donald Trump en 2016, Cambridge Analytica intervino en más de sesenta procesos electorales en todo el mundo. Su modelo era replicable: segmentación psicográfica precisa, mensajes emocionales personalizados y explotación de vulnerabilidades sociales.

Durante el referéndum del Brexit, colaboraron con el grupo Leave.EU enfocándose en votantes indecisos mediante mensajes que apelaban al miedo a la inmigración o a la pérdida de identidad nacional. En la campaña de Trump, se crearon más de 50 000 versiones diferentes de anuncios adaptados a microsegmentos y se desplegaron campañas específicas para desalentar a ciertos simpatizantes demócratas de votar.

Su alcance no se limitó a democracias consolidadas: también operaron en Kenia, México, India o Nigeria. En Kenia, por ejemplo, los mensajes polarizantes buscaron reforzar divisiones étnicas con fines electorales. Incluso en contextos como estos, tan diversos, la lógica conductual pudo ser aplicada eficazmente.

Consecuencias y escándalo global

El escándalo estalló en 2018 cuando Christopher Wylie, exanalista de la empresa, reveló su funcionamiento interno. Gracias a investigaciones de *The Guardian*, por Cadwalladr y Graham-Harrison (2018), y de *The New York Times*, la opinión pública descubrió la magnitud de esta operación: una estrategia sistemática de manipulación psicológica basada en datos personales.

Facebook fue multada con 5000 millones de dólares por la Comisión Federal de Comercio de Estados Unidos y Cambridge Analytica cerró sus puertas. Pero más allá de las sanciones, el caso dejó al descubierto un fenómeno más amplio: el uso de tecnologías predictivas y psicográficas como herramientas operativas en procesos de decisión política.

Desde mi perspectiva, formada en el análisis del comportamiento y la negociación en contextos complejos, lo más revelador no fue solo su dimensión tecnológica, sino la estructura que lo sostenía. Una estructura progresiva, calculada, centrada en la identificación de patrones emocionales y en la adaptación estratégica del mensaje para generar una respuesta específica.

Convergencias con la negociación conductual

Al examinar en detalle los mecanismos descritos hasta aquí, desde la recopilación de datos hasta la personalización del mensaje y la búsqueda de una reacción conductual, resulta difícil no reconocer un paralelismo con ciertos modelos utilizados en negociación profesional.

No se trata aquí de sugerir que toda forma de persuasión digital sea equivalente a una táctica de negociación. Pero sí de observar que muchos de los principios aplicados en campañas de *microtargeting* emocional, escucha pasiva, lectura emocional, calibración progresiva de mensajes o generación de respuesta se asemejan estructuralmente a lo que implementamos en intervenciones humanas complejas.

En mi experiencia, influir sobre el comportamiento de otro no es cuestión de presión o repetición, sino de secuencia, empatía y ajuste. Precisamente por eso, me parece útil y revelador analizar un modelo que sintetiza esta lógica: el modelo de la escalera de cambio conductual (inglés, *behavioral change stairway model*), desarrollada por el FBI para situaciones de crisis.

Sin perder de vista el contexto para el cual fue diseñado, en la sección siguiente propongo examinar cómo este modelo de negociación interpersonal puede ofrecernos claves reveladoras para comprender la lógica que guía muchas de las estrategias persuasivas actuales, incluidas aquellas implementadas por sistemas digitales.

NEGOCIACIÓN E INFLUENCIA: EL MODELO CONDUCTUAL DEL FBI

Las técnicas de negociación han sido históricamente empleadas en situaciones álgidas: desde crisis de rehenes hasta contextos diplomáticos altamente sensibles donde la palabra puede ser literalmente la diferencia entre la vida y la muerte. Entre los modelos más reconocidos, destaca la escalera de cambio conductual, desarrollada por el FBI para guiar procesos de cambio de comportamiento en contextos de máxima tensión (Vecchi, Hasselt & Romano, 2005).

Este enfoque plantea una secuencia de cinco etapas: escucha activa, empatía, conexión, influencia y, finalmente, transformación conductual. Cada fase requiere tiempo, atención genuina y una lectura fina del estado emocional del interlocutor.

Lo llamativo, desde la perspectiva actual, es que muchos de estos principios, originalmente pensados para la interacción humana directa, han sido reinterpretados en el entorno digital. Hoy, sistemas algorítmicos que emplean inteligencia artificial simulan empatía, personalizan el discurso y buscan generar una respuesta conductual adaptada. A simple vista, podría parecer una extrapolación arriesgada, pero, al observar de cerca cómo funcionan ciertas plataformas digitales, la analogía se vuelve difícil de ignorar.

De la negociación humana a la automatización algorítmica

En la era digital, la lógica secuencial del modelo del FBI —escuchar, empatizar, conectar, influir y transformar— no ha desaparecido, simplemente, ha cambiado de rostro. Hoy, esa misma arquitectura opera en plataformas digitales que analizan nuestras huellas emocionales y adaptan sus respuestas con una precisión sin precedentes.

En el entorno digital contemporáneo, cada etapa del modelo del FBI encuentra un correlato funcional en la interacción entre usuarios y algoritmos:

Tabla 1

Etapas del modelo del FBI y su equivalencia en entornos digitales

Etapa del modelo del FBI	Equivalente en entornos digitales
Escucha activa	Recojo pasivo de datos digitales (clics, vistas, interacciones)
Empatía	Análisis de sentimiento y emoción en lenguaje natural mediante *natural language processing* (NLP) (español, procesamiento de lenguaje natural)
Conexión	Personalización algorítmica de contenidos según valores e intereses
Influencia	*Microtargeting* conductual y emocional
Cambio de comportamiento	Conversión medible: clics, votos, compras, opiniones, etc.

Mientras que los negociadores del FBI necesitaban horas, a veces días enteros, para construir una mínima base de confianza con su interlocutor, los algoritmos actuales logran detectar patrones emocionales y conductuales en apenas unos segundos de interacción. El contraste es brutal. Aplicaciones como TikTok, por ejemplo, no solo identifican con precisión qué tipo de contenido genera placer, sorpresa o molestia: también ajustan al instante el flujo de videos para mantenernos enganchados sin que lo notemos. Lo que en la negociación tradicional requería presencia, intuición y paciencia, hoy se replica y se acelera mediante procesos automáticos, invisibles y altamente eficaces.

Técnicas clásicas reapropiadas por la IA

Varios recursos clave de la negociación persuasiva han sido absorbidos y transformados por la inteligencia artificial. No se trata solo de una copia funcional, sino de una reinterpretación adaptada a entornos digitales de alta velocidad y bajo umbral de atención:

- La presencia estratégica ha sido sustituida por interfaces amables, asistentes virtuales o *chatbots* que imitan atención genuina, aunque detrás no haya nadie.
- La reformulación se hace mediante los sistemas de recomendación que ajustan los mensajes según el historial del usuario, como si reencuadraran silenciosamente la conversación, una técnica muy conocida entre negociadores humanos.

- La atención emocional se realiza mediante procesamiento del lenguaje natural (NLP), modelos como *bidirectional encoder representations from transformers* (BERT) o *generative pre-trained transformer* (GPT) analizan las emociones implícitas en el texto y adaptan las respuestas a los estados de ánimo de forma casi instantánea (Colneric & Demsar, 2020).

Estos paralelismos no son anecdóticos, sino que responden a una lógica calculada de optimización de la influencia. Así como un negociador experto busca crear un terreno común con su interlocutor, las plataformas digitales construyen una ilusión de personalización ofreciendo al usuario algo que parece hecho a su medida, aunque en realidad se trate de una estrategia cuidadosamente orientada a fines comerciales o políticos. Precisamente, ese desfase entre lo que percibimos y lo que realmente ocurre es lo que plantea una de las tensiones más delicadas de esta nueva forma de persuasión: la ilusión de autonomía frente a una influencia cuidadosamente orquestada.

El riesgo de la negociación invisible

En las interacciones tradicionales, la persuasión se presenta de forma clara: uno sabe cuándo está frente a un político, un vendedor o un negociador. Hay códigos, roles, expectativas y, por lo tanto, también hay defensa: el receptor activa filtros cognitivos, evalúa, acepta o rechaza el intento de influencia.

En cambio, cuando la inteligencia artificial entra en juego, las fronteras tradicionales de la persuasión se diluyen. No hay rostro ni tono ni advertencia, solo una interfaz aparentemente neutral que recomienda, sugiere o adapta, pero no se declara. El resultado es que muchos usuarios no perciben que están siendo persuadidos y esa falta de percepción es justamente lo que debilita su capacidad de reflexión y resistencia crítica.

Da Empoli (2024) sostiene que los nuevos actores políticos digitales ya no buscan convencer racionalmente al electorado, sino activar emociones primarias capaces de movilizar comportamientos inmediatos. Como señala Sunstein (2017), estas nuevas formas de influencia no buscan imponer, sino empujar suavemente (inglés, *nudge*) al sujeto hacia una dirección determinada. Es una negociación sin rostro, sin nombre y, muchas veces, sin conciencia. El consentimiento no desaparece, pero se vuelve irrelevante: lo sustituye la inercia.

IMPACTOS SOCIALES Y ÉTICOS: RIESGOS EN LA ERA DE LA PERSUASIÓN INVISIBLE

La integración creciente de sistemas de inteligencia artificial en procesos comunicativos, políticos y comerciales ha abierto un universo de posibilidades en cuanto a personalización, segmentación y eficacia del mensaje. Sin embargo, junto con estas oportunidades, emergen zonas grises, a veces francamente oscuras, que exigen una revisión ética urgente.

El caso de Cambridge Analytica fue un punto de inflexión que no solo reveló el potencial de estas tecnologías, sino también el vacío legal que las rodea, los límites difusos entre persuasión e

interferencia, y el riesgo real de que la ciudadanía se convierta en blanco de estrategias de manipulación tan sofisticadas que operan por debajo del umbral de la conciencia.

Manipulación sin consentimiento

El riesgo más delicado quizá no sea la mentira directa ni la censura, sino la posibilidad de ser guiado sin saberlo, la capacidad del algoritmo de anticipar deseos, temores o decisiones futuras mediante el análisis masivo de datos. Lo que Zuboff (2019) denomina excedente conductual, el cual permite diseñar mensajes que no solo apelan a lo que pensamos hoy, sino que influyen en lo que pensaremos mañana.

En este punto, la frontera entre persuasión legítima y manipulación se vuelve borrosa. Cuando el receptor no sabe que está siendo dirigido, la posibilidad de una respuesta crítica informada se disuelve. Sunstein (2017) lo advierte claramente: estos "empujones" digitales alteran la arquitectura misma de nuestras decisiones sin necesidad de coerción ni promesas explícitas. ¿Es esto siempre negativo? No necesariamente.

Algunas de estas técnicas pueden servir para promover conductas saludables o prevenir comportamientos de riesgo. Pero cuando se aplican sin consentimiento o sin supervisión ética —cuando quien empuja no busca el bienestar del usuario, sino su explotación política, económica o emocional—, estamos ante una amenaza directa a la autonomía personal. Y ese riesgo no es hipotético: ya está entre nosotros.

Sesgos algorítmicos y discriminación automatizada

Los sistemas de inteligencia artificial no piensan por cuenta propia: aprenden a partir de enormes volúmenes de datos que les proporcionamos. Si esos datos están impregnados de prejuicios históricos, desigualdades estructurales o estereotipos culturales —como sucede con frecuencia—, los algoritmos no los corrigen; todo lo contrario, los replican y, en muchos casos, los profundizan.

Esto implica que, pese a su apariencia técnica, los algoritmos no son imparciales ni objetivos, sino que. reflejan las decisiones, creencias y limitaciones de quienes los diseñan, y reproducen los sesgos que ya existen en nuestras sociedades. Como advierte Eubanks (2018), la inteligencia artificial puede convertir antiguas formas de discriminación en sistemas automáticos, opacos y muy difíciles de detectar.

En los ámbitos político y comunicacional, estas distorsiones pueden tener consecuencias muy serias. Por ejemplo, si se utiliza un sistema de análisis de sentimientos para interpretar discursos en redes sociales, este podría etiquetar como agresivo o negativo un mensaje que, en realidad, expresa una denuncia legítima o una forma de resistencia, especialmente si proviene de comunidades históricamente excluidas. Como señala Noble (2018), los algoritmos no solo organizan la información, también deciden qué voces se amplifican y cuáles quedan silenciadas.

El problema se agrava aún más cuando estas técnicas se aplican al terreno electoral. Tras el escándalo de Cambridge Analytica, se descubrió que en Estados Unidos algunos grupos de población

afroamericana fueron objetivo de campañas específicas destinadas a desmotivarlos para que no acudieran a votar. Estos mensajes, diseñados para generar apatía, explotaban emociones como la frustración, la desconfianza en las instituciones o el desencanto político sin siquiera recurrir a métodos de exclusión manifiestos.

Esta manipulación automatizada, casi imposible de rastrear, representa una forma especialmente insidiosa de injusticia social. No actúa a través de leyes ni discursos públicos, sino desde los márgenes invisibles del código, los datos y los algoritmos. Precisamente por eso es tan difícil de enfrentar.

Desinformación, polarización y *deepfakes*

La capacidad de generar contenido falso con alto poder persuasivo, como imágenes o videos sintéticos conocidos como *deepfakes*, constituye otra amenaza de peso. Estas tecnologías permiten difamar, distorsionar hechos o confundir al electorado debilitando la confianza en los medios, actores políticos e instituciones democráticas.

La lógica misma de los algoritmos de recomendación, que priorizan contenidos con alto potencial de interacción, favorece la polarización emocional. Los mensajes más extremos, alarmistas o moralmente cargados tienden a generar más clics, más comentarios, más circulación. En ese entorno, muchos usuarios terminan atrapados en burbujas informativas donde solo se refuerzan sus ideas previas; con ello, se pierde la posibilidad del diálogo y de la escucha mutua (Sunstein, 2017).

Erosión de la confianza pública

Quizás el efecto más profundo y a la vez más insidioso del uso irresponsable de estas tecnologías sea la erosión progresiva de la confianza en la información que recibimos, en los procesos democráticos que estructuran nuestras sociedades y en la veracidad misma de los discursos públicos.

Cuando las personas descubren que sus opiniones pueden haber sido manipuladas, que los contenidos que consumen no son neutros, y que sus propios datos se usan para anticipar y modificar su conducta, el resultado es un clima generalizado de sospecha que va carcomiendo el tejido social.

Este fenómeno es especialmente dañino en las democracias, donde el debate público se basa en la posibilidad de deliberar libremente, contrastar argumentos y tomar decisiones informadas. Sin confianza, todo ese andamiaje tambalea, incluida la legitimidad misma del sistema democrático.

DISCUSIÓN: ENTRE AUTONOMÍA, PODER Y REGULACIÓN

Los avances tecnológicos aplicados a la comunicación y a la persuasión nos enfrentan a una paradoja difícil de ignorar: nunca habíamos tenido tanto acceso a la información y, sin embargo, nunca fue tan fácil moldear nuestras opiniones sin que siquiera lo notemos. La combinación de inteligencia artificial, análisis de datos conductuales y modelos psicológicos ha desplazado el poder persuasivo desde el mensaje hacia la ingeniería emocional y cognitiva del receptor. Ya no basta con decir lo adecuado; ahora se trata de decirlo a la persona indicada, en el momento exacto, con el tono justo y la carga emocional precisa.

Este escenario plantea una discusión urgente sobre los límites entre influencia legítima y manipulación encubierta. También nos obliga a repensar el rol que deben asumir los distintos actores —Estados, empresas, ciudadanía y academia— para resguardar pilares como la autonomía individual, la deliberación informada y el bien común.

¿Es posible una persuasión ética basada en datos?

Toda forma de persuasión implica influir en otro. Eso no es nuevo ni necesariamente problemático. El problema ético aparece cuando se diluyen tres elementos clave: la transparencia del proceso, el consentimiento del receptor y su capacidad de reflexionar críticamente sobre lo que recibe. El problema no radica en el uso de datos en sí, sino en su utilización opaca: cuando el ciudadano desconoce cómo, por qué y con qué finalidad sus datos son empleados para construir perfiles y dirigir mensajes personalizados.Algunos expertos han propuesto principios éticos aplicables al diseño de algoritmos persuasivos: trazabilidad de las decisiones automatizadas, explicabilidad de los procesos y no discriminación (Floridi et al., 2018). También destacan que los usuarios deberían saber cuándo están siendo influidos por sistemas de IA, y poder ejercer un control real sobre sus datos y sobre las formas en que acceden a la información.

La urgencia de marcos regulatorios

Uno de los efectos más visibles del caso Cambridge Analytica fue mostrar, sin matices, el vacío legal que rodea el uso de datos personales con fines políticos. En muchos países, las leyes vigentes de protección de datos son simplemente insuficientes frente a la complejidad del análisis psicográfico y de la microsegmentación automatizada.

Algunas iniciativas, como el Reglamento General de Protección de Datos (GDPR) (2016) en Europa, han marcado avances importantes al introducir principios como el consentimiento informado, el derecho al olvido o la portabilidad de datos. Sin embargo, aún estamos lejos de contar con normativas específicas y eficaces para regular el uso de IA en campañas electorales, publicidad política y plataformas digitales. Especialmente, en contextos donde las instituciones democráticas son frágiles o están en disputa.

Educación crítica y resiliencia ciudadana

Más allá de las políticas públicas y los marcos legales, es indispensable fortalecer la capacidad crítica de la ciudadanía. Saber cómo funcionan los algoritmos, reconocer los sesgos cognitivos que nos hacen más vulnerables a ciertas formas de persuasión y desarrollar herramientas para verificar la información que consumimos, se ha convertido en una competencia cívica básica del siglo XXI.

La alfabetización mediática, la ética informacional y la promoción del pensamiento crítico deberían ocupar un lugar central en los programas educativos. Una ciudadanía informada, crítica y consciente es, probablemente, la defensa más eficaz frente a formas de influencia que operan desde la invisibilidad.

Un nuevo contrato entre tecnología y democracia

Todo lo anterior nos lleva a imaginar un nuevo contrato social entre tecnología y democracia. Uno que no reniegue del potencial transformador de la inteligencia artificial, pero que le imponga límites claros, garantice derechos fundamentales y promueva una cultura del cuidado digital.

La persuasión no va a desaparecer, ni debe hacerlo, pero sí necesita volver a enraizarse en principios éticos, en valores democráticos, y en el respeto por la libertad individual y el interés colectivo.

CONCLUSIONES

Esta investigación me ha permitido constatar, con claridad más que con alarma, que la combinación entre recolección masiva de datos digitales, inteligencia artificial, psicografía y modelos de negociación conductual ha transformado de forma profunda y silenciosa los mecanismos mediante los cuales se configura la opinión pública en el entorno digital contemporáneo. A diferencia de los medios tradicionales, donde la persuasión era visible, deliberada y generalmente masiva, hoy asistimos a formas de influencia que operan en lo íntimo, lo personalizado y lo automatizado.

Casos como el de Cambridge Analytica no son simples excesos, sino síntomas de un cambio de época. Han demostrado que los algoritmos no solo organizan la información, sino que actúan como intermediarios políticos silenciosos capaces de identificar motivaciones emocionales, adaptar mensajes y generar comportamientos específicos sin que el usuario sea plenamente consciente del proceso. Estas prácticas se apoyan en técnicas de perfilamiento psicográfico, como el modelo OCEAN, y en principios derivados de modelos de negociación conductual, especialmente el desarrollado por el FBI para contextos de crisis.

Conozco bien ese modelo, la escalera de cambio conductual, porque lo he aplicado en situaciones reales, donde escuchar y conectar puede literalmente salvar vidas. Por eso me resulta tan revelador ver cómo sus cinco etapas (escucha activa, empatía, conexión, influencia y cambio de comportamiento) han sido trasladadas al entorno digital no por negociadores entrenados, sino por sistemas algorítmicos. Hoy, cada una de esas fases puede ser activada por datos, procesada en tiempo real y puesta al servicio de fines comerciales, políticos o ideológicos.

El problema no es solo técnico: es también ético. Esta forma de negociación algorítmica se despliega sin transparencia, sin consentimiento informado y, con frecuencia, sin rendición de cuentas. La empatía simulada puede ser funcional, pero nunca es genuina; y las decisiones que induce no surgen de una deliberación libre, sino de una arquitectura persuasiva invisible. En este nuevo ecosistema, la inteligencia artificial no solo redibuja la comunicación política: reformula las condiciones mismas de la autonomía individual y del debate democrático.

Entre los riesgos más relevantes identificados en esta investigación, destaco cuatro: la manipulación imperceptible, la reproducción automatizada de sesgos estructurales, la diseminación personalizada de desinformación y la fragmentación progresiva del espacio público. Ante los riesgos identificados, esta investigación plantea una respuesta articulada en tres niveles:

- Ético: es necesario establecer límites claros —alineados con principios de justicia, responsabilidad y respeto por la dignidad humana— al diseño y uso de sistemas persuasivos basados en IA.
- Regulativo: se requieren marcos normativos específicos que reconozcan el poder real de estas tecnologías y protejan a la ciudadanía; especialmente, en ámbitos sensibles como campañas electorales, decisiones públicas automatizadas o entornos judiciales.
- Educativo: no basta con saber usar estas tecnologías, es imprescindible comprender cómo funcionan, cuándo estamos siendo influidos y cómo recuperar capacidad crítica y agencia. Esto incluye desarrollar una alfabetización en negociación digital capaz de detectar cuándo una sugerencia amistosa es en realidad una estrategia cuidadosamente calibrada.

En definitiva, el desafío que enfrentamos no es solo tecnológico: es cultural y profundamente político. La inteligencia artificial, como toda herramienta poderosa, puede fortalecer nuestras democracias o erosionarlas desde dentro. No se trata de rechazar el avance, sino de orientar ese avance hacia el interés público, el diálogo informado y el respeto a las personas como sujetos de derecho, no como objetos de manipulación.

La libertad, lo creo firmemente, no se defiende desconectándose, sino entendiendo cómo no cederla sin darnos cuenta. En una época de negociaciones invisibles entre humanos y algoritmos, preservarla exige más que tecnología: requiere conciencia, criterio y voluntad de respuesta.

REFERENCIAS

Aristóteles. (1999). *Retórica* (A. Tovar, Ed.). Centro de Estudios Políticos y Constitucionales. (Obra original publicada ca. 350 a. C.)

Cadwalladr, C., & Graham-Harrison, E. (2018, 17 de marzo). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian.* https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

Cialdini, R. B. (2007). *Influence: The psychology of persuasion* (Edición revisada). Harper Business.

Colneric, N., & Demsar, J. (2020). Emotion recognition on Twitter: Comparative study and training a unison model. *IEEE Transactions on Affective Computing, 13*(1), 402-416.

Da Empoli, G. (2024). *Les ingenieurs du chaos* (N. Boullosa, Trad.). OBERON. (Obra original publicada en 2019)

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor.* St. Martin's Press.

Floridi, L., Cowls, J., Beltrametti, M., Chiarello, F., Chatila, R., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707.

Jamieson, K. H., & Campbell, K. K. (2001). *The interplay of influence: News, advertising, politics, and the mass media* (5.ª ed.). Wadsworth.

Jowett, G. S., & O'Donnell, V. (2018). *Propaganda and persuasion* (7.ª ed.). Sage.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kennedy, G. A. (2007). *Una historia de la retórica clásica*. Gredos.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, 110*(15), 5802-5805.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences, 114*(48), 12714-12719.

McLuhan, M. (1964). *Understanding media: The extensions of man.* MIT Press.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press.

Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business.* Viking Penguin.

Reglamento general de protección de datos, 679, Diario Oficial de la Unión Europea (2016). https://www.boe.es/doue/2016/119/L00001-00088.pdf

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media.* Princeton University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

Vecchi, G. M., Hasselt, V. B. van., & Romano, S. J. (2005). Crisis (hostage) negotiation: Current strategies and issues in high-risk conflict resolution. *Aggression and Violent Behavior, 10*(5), 533–551.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* PublicAffairs.